Title: Evidence for abstract codes in parietal cortex guiding prospective working memory

Jongmin Lee[1*], David De Vito[1], Jacob A. Miller[2] & Derek Evan Nee[1]

[1] Department of Psychology, Florida State University, Tallahassee, FL, USA

[2] Department of Psychology, University of Miami, Coral Gables, FL, USA

[*] Correspondence: jongmin.lee.529@gmail.com

Abstract

The recent past helps us predict and prepare for the near future. Such preparation relies on working memory (WM) which actively maintains and manipulates information providing a temporal bridge. Numerous studies have shown that recently presented visual stimuli can be decoded from fMRI signals in visual cortex (VC) and the intra-parietal sulcus (IPS) suggesting that these areas sustain the recent past. Yet, decoding concrete, sensory signals leaves open how past information is transformed into the abstract codes critical for guiding future cognition. Here, human participants used WM to maintain a separate spatial location in each hemifield wherein locations were embedded in a star-shaped sequence. On each trial, participants made a sequence-match decision to a spatial probe and then updated their WM with the probe. The same abstract star-shaped sequence guided judgments in each hemifield allowing us to separately track concrete spatial locations (hemifield-specific) and abstract sequence positions (hemifield-general), while also tracking representation of the past (last location/position) and future (next location/position). Consistent with previous reports, concrete past locations could be decoded from VC and IPS. Moreover, in anticipation of the probe, representations shifted from past to future locations in both areas. Critically, we observed abstract coding of future sequence positions in the IPS whose magnitude related to speeded performance. These data suggest that the IPS sustains abstract codes to facilitate future preparation and reveal the transformation of the sensory past into the abstract future.

Keywords

working memory, fMRI, parietal cortex, intra-parietal sulcus, visual cortex, abstraction, decoding

Introduction

Imagine performing instructed boxing combinations such as jab twice with the lead hand, hook with the other, then uppercut with the lead hand. To do so the combination needs to be held in mind with each strike chaining into the next. This task requires working memory (WM), which refers to the maintenance and manipulation of information guiding ongoing behaviors [1-4]. WM is central to much of higher-level cognition as many of our behaviors require extending the recent sensory past in order to properly act in the moment and near future.

Much of our use of WM is prospective in nature [5-8] as the recent past allows us to prepare for the near future. Yet, most studies focus on how the recent past is maintained. Standard paradigms present participants with sensory information such as colors, orientations, spatial locations, or motion directions and then ask participants to recognize or reproduce the previously presented stimuli. A wealth of studies using such paradigms have demonstrated that high-fidelity stimulus-specific codes can be decoded from fMRI activation patterns in visual and parietal cortices [9-21]. Findings such as these have given rise to the sensory recruitment hypothesis which states that WM is supported by the sustained maintenance of sensory information in those same sensory cortices involved in initial perception [22-25].

While substantial sensory information is held in mind to support WM, much of our WM is guided by non-sensory codes. In particular, many of our prospective actions are guided by more abstract codes. For example, we can apply the same boxing combination described above when in an orthodox (left foot/hand forward) or southpaw stance (right foot/hand forward) such that the hand performing each strike is reversed as a function of stance. Abstract coding affords generalization which facilitates adaptive behavior and knowledge transfer across contexts [26]. Despite the importance of generalization to WM and cognition, few studies have examined abstract coding in support of WM. Some evidence has demonstrated that concrete sensory information of the past can be recoded into a more abstract format such as translating a motion cloud to a line [27, 28]. Moreover, some recent studies have demonstrated that when forthcoming actions can be anticipated,

prospective action codes are maintained [29-32]. However, there is a critical gap in our understanding of the nature and process of the abstract representations that reside in-between sensation and action which prospectively and flexibly guide WM for the future.

Abstract coding increases with synaptic distance from primary cortices [33]. For visuo-spatial processing, the intra-parietal sulcus (IPS) is downstream of visual cortical areas of the occipital cortex (e.g., V1-V4; VC). Although in numerous cases, visual WM is decoded similarly in VC and IPS [12, 19-21] some work has demonstrated that IPS WM representations are more robust to distraction [10, 16, 17]. It has been speculated that abstraction results in a stable, low dimensional coding format that is distractor resistant [17, 34]. Hence, we reasoned that abstract codes that facilitate prospective uses of WM would be relatively more prominent in the IPS than VC.

The current study used fMRI to investigate the mechanisms underlying the use of the recent past to prepare for the future. Human participants performed a sequence-matching task which probed whether a presented spatial location followed the last spatial location in a predefined sequence (**Fig 1**). The sequence (**Fig 1a**) was abstract in that it could be generalized to apply to spatial locations in either hemifield. That is, we define abstraction as the ability to generalize over contexts or instances [27, 28, 35-38]. Moreover, the use of a sequence afforded the use of the recent past (most recently presented location) to predict and prepare for the near future (next location in the sequence). Through this task, we examined concrete (spatial location), abstract (sequence position), retrospective (previous location/position), and prospective (next location/position) coding in visual and parietal cortices. We predicted that retrospective and concrete coding would dominate VC consistent with the sensory recruitment hypothesis. Moreover, we anticipated more abstract and prospective coding in the IPS consistent with a top-down code to guide future expectations. Such a result would provide evidence for the use of abstract codes to guide prospective WM.

Results

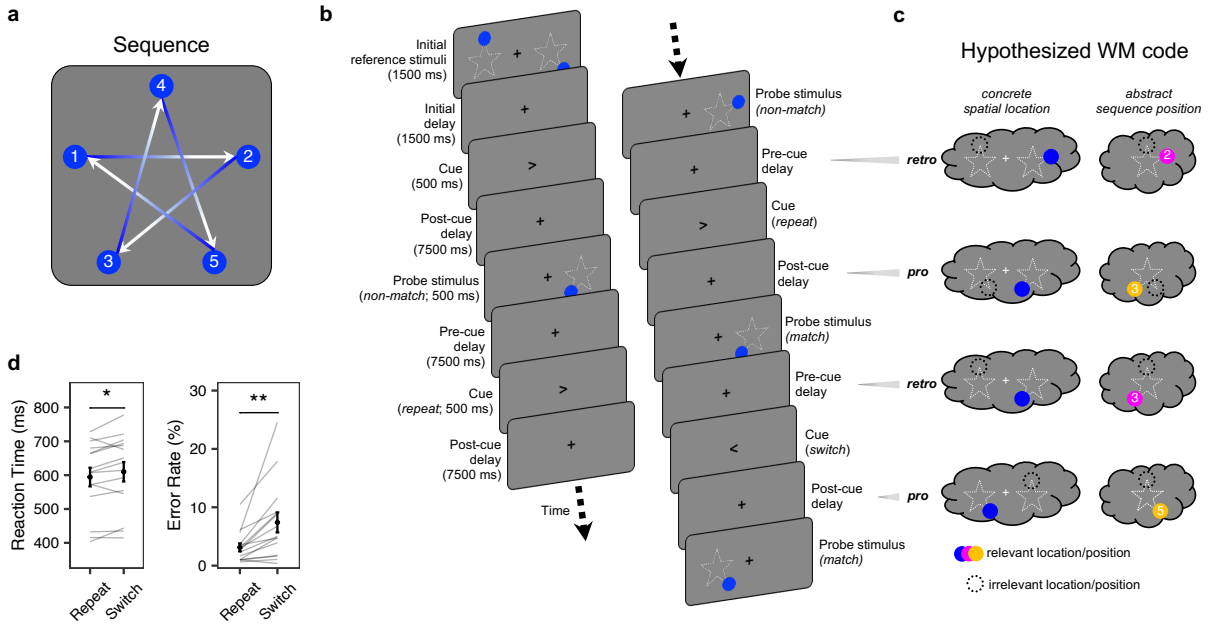

Fig 1. Procedures and behavioral performance.
(a) The sequence consisted of five points forming a star. Arrows indicate the sequence ordering such that position 2 followed position 1, position 3 followed position 2, etc. Position 5 transitioned back to position 1 such that the sequence was circular.
(b) Task procedure. Each trial consisted of a pre-cue delay, a cue, and a post-cue delay. At the start of each run, initial spatial locations were presented in the left and right visual hemifields. After an initial delay, a cue indicated the task-relevant visual hemifield wherein a forthcoming probe would be presented following a post-cue delay. Participants determined whether the spatial location of the probe followed the spatial location of the last item that appeared in the same hemifield. Half of the probes were sequence matches (e.g., position 2 following position 1) while half were non-matches (e.g., position 3 following position 1). The star trace is used for illustrative purpose and did not appear in the actual task.
(c) We hypothesized that during the pre-cue delay, the spatial location of the just-probed location (concrete, retrospective code; hemifield-specific) and its sequence position (abstract, retrospective code; hemifield-general) would be maintained in working memory. During the post-cue delay, we hypothesized that retrospective codes would be transformed towards future expectations. Specifically, the next sequence position (abstract, prospective code; hemifield-general) would lead to a prediction of the next spatial location in the sequence (concrete, prospective code; hemifield-specific). For completeness, we depict analogous coding possibilities for the irrelevant (uncued) locations/positions although these were not of central interest.
(d) Mean reaction time and error rates. Behavioral performance was better when the cue indicated the same visual hemifield as the previous trial ('Repeat') relative to when the cue indicated the other visual hemifield ('Switch'). Each line on the bar graph represents each participant. Bold dots indicate the mean across participants with error bars representing ± 1 standard error of mean. *$p < 0.05$, **$p < 0.01$.

Participants separately and continuously tracked spatial locations in the left and right visual hemifields (**Fig 1b**). On each trial, a cue indicated the relevant hemifield for the forthcoming probe affording a prospective prediction of the sequencing-matching location. Participants indicated whether the spatial location of the probe followed the spatial location of the last stimulus in the cued hemifield (reference). Half of the probes were matches while half were non-matches with non-match probes equally drawn from the three non-matching spatial locations excluding the reference (i.e., there were no stimulus repeats). After responding, the spatial location of the probe became the reference for the next probe of the same hemifield. The cue was separated from the previous (pre-cue delay) and forthcoming (post-cue delay) probes allowing for periods to track WM codes of different forms (retrospective vs prospective, concrete/location vs abstract/position; **Fig 1c**).

Behavioral data indicated that participants performed the task well (mean accuracy = 94.8 %). To verify that participants appropriately prioritized relevant information (i.e., the cued hemifield), we separately characterized trials in which the cue indicated the same visual hemifield as the previous trial (‘Repeat’) or the opposite visual hemifield (‘Switch’). As anticipated, reaction times (RT) and error rates (ER) were larger in ‘Switch’ (610.1 ms, 7.4 %) relative to ‘Repeat’ trials (594.5 ms, 3.1 %; $t_{(14)} = 2.53$, $p < .05$ for reaction time; $t_{(14)} = 3.44$, $p < .01$ for error rate; **Fig 1d**) consistent with a cost for switching priority. These data confirm that participants performed the task as expected.

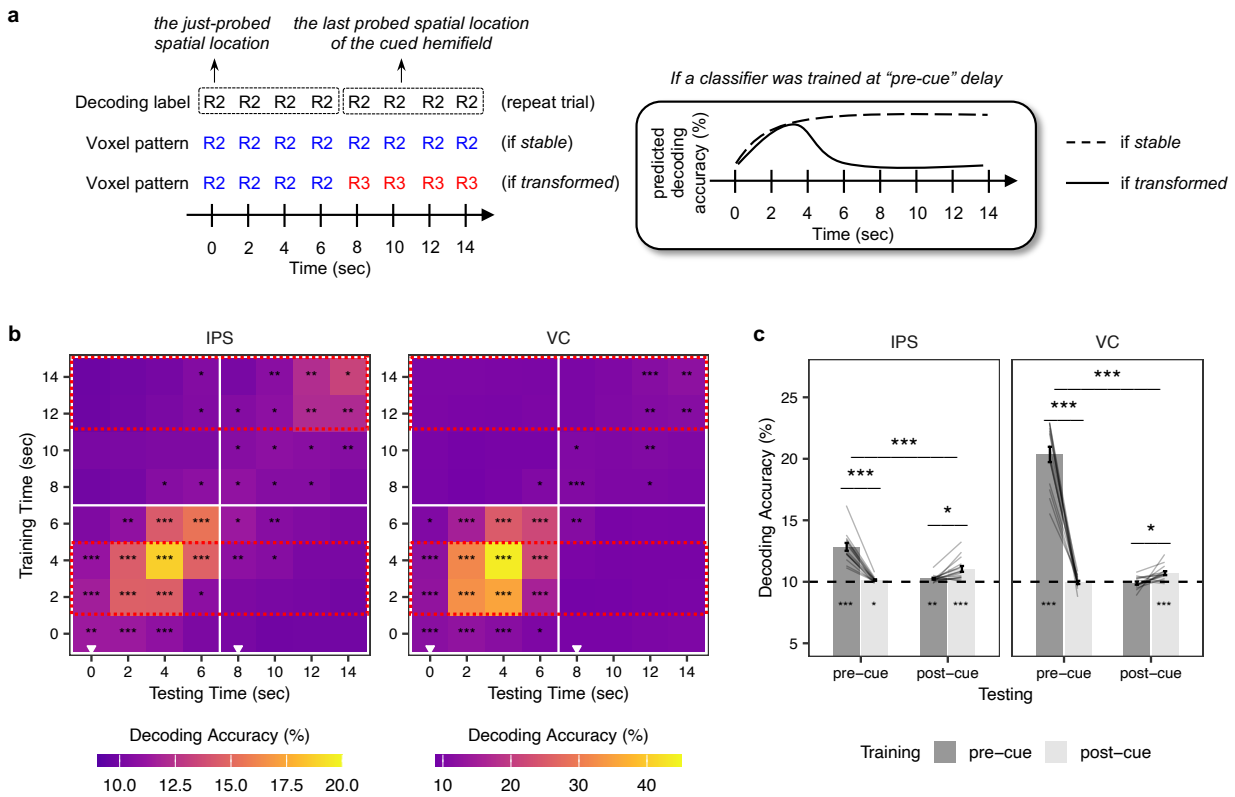

Fig 2. Working memory content maintained during the post-cue delay differs from that maintained during the pre-cue delay.
(a) Each time point was labeled based on the just-probed location (pre-cue) and/or last probed location in the cued hemifield (post-cue). Whether the WM remained stable or was transformed predicts different patterns of decoding accuracy across pre- and post-cue delay periods.
(b) Temporal generalization decoding. Classifiers were trained on a given time point and tested across time points. Decoding performance was compared against the chance level of 10% and corrected for multiple comparisons using false discovery rate (FDR). The red dotted boxes on the graph indicate the time points of maximal decoding for a given period which were used for training in subsequent analyses (**Fig 3**). *q < 0.05 and **q < 0.01 (FDR corrected).
(c) Average of decoding accuracies in each delay period. The horizontal dotted line indicates the chance level of 10%. Each line on the bar graph represents each participant. *p < 0.05, **p < 0.01, ***p = 0.001. Error bars represent ±1 standard error of mean.

Concrete location codes are transformed over time.

We hypothesized that WM codes would transform over time trading off from retrospective to prospective, and concrete to abstract (**Fig 1c**). To test for representational transformations, we performed temporal generalization decoding [39]. For each trial, time points (TRs) were labeled by their reference location in visual space (i.e., the just-probed

spatial location during the pre-cue delay; the last probed spatial location of the cued hemifield during the post-cue delay; **Fig 2a**). Separate classifiers were trained for each time point and region-of-interest (ROI) on fMRI activation patterns distinguishing each of the 10 spatial locations. ROIs included the left and right IPS and VC. Training and testing were performed separately for each hemifield (**Fig S1**), and then results were averaged across hemifields for each of the IPS and VC, respectively (**Fig 2b**). If WM coding is stable, classifications should generalize across time points (i.e., significant off-diagonal decoding across delays). If WM coding is fully dynamic, classifications should not generalize across time points (i.e., no off-diagonal decoding). However, if distinct WM codes underlie each of the pre- and post-cue periods, classifiers trained during each period should generalize within a delay period, but not generalize across delay periods (i.e., off-diagonal decoding restricted to same delay; **Fig 2b**). The results were most consistent with this latter hypothesis. Decoding was consistently above chance when training and testing were carried out in the same delay period (number of significant decoding time points in left VC: 29/32, right VC: 22/32, averaged VC: 24/32, left IPS: 28/32, right IPS: 25/32, averaged IPS: 28/32) but consistently near chance levels when the training and testing were carried out in different delay period (number of significant decoding time points in left VC: 2/32, right VC: 3/32, averaged VC: 2/32, left IPS: 9/32, right IPS: 3/32, averaged IPS: 8/32; **Fig S1**; **Fig 2b**).

To further quantify this result, we averaged the decoding accuracies within the delay periods (from 0 to 6 sec, ‘pre-cue’; from 8 to 14 sec, ‘post-cue’) in each quadrant of the temporal generalization decoding matrix (**Fig 2c**). This resulted in 2 (training period: ‘pre-cue’ vs ‘post-cue’) x 2 (testing period: ‘pre-cue’ vs ‘post-cue’) sets of decoding accuracies for each of the IPS and VC which were submitted to two-way repeated-measures analyses of variance (RM ANOVA) with training period (‘pre-cue’ vs. ‘post-cue’) and testing period (‘pre-cue’ vs. ‘post-cue’) as within-subject factors. In both the IPS and VC, we observed a significant main effect of training period with superior performance when training on the ‘pre-cue’ period (IPS: $F_{(1, 14)} = 93.95$, $p = .001$; VC: $F_{(1, 14)} = 295.15$, $p = .001$). Similarly, we observed a significant main effect of testing time with superior performance when testing the ‘pre-cue’ period (IPS: $F_{(1, 14)} = 101.61$, $p = .001$; VC: $F_{(1, 14)}$

= 315.43, p = .001). Critically, in both areas, a significant interaction was observed such that performance was superior when training and testing were performed in the same period relative to when they differed (IPS: $F_{(1, 14)} = 57.92$, p = .001; VC: $F_{(1, 14)} = 167.99$, p = .001). Such data support the idea that distinct representational codes were maintained during the pre- and post-cue delay periods. However, since representations of the same stimuli can transform due to prioritization and selection processes [40-43], this in-and-of-itself does not demonstrate that the post-cue delay period contains a prospective code.

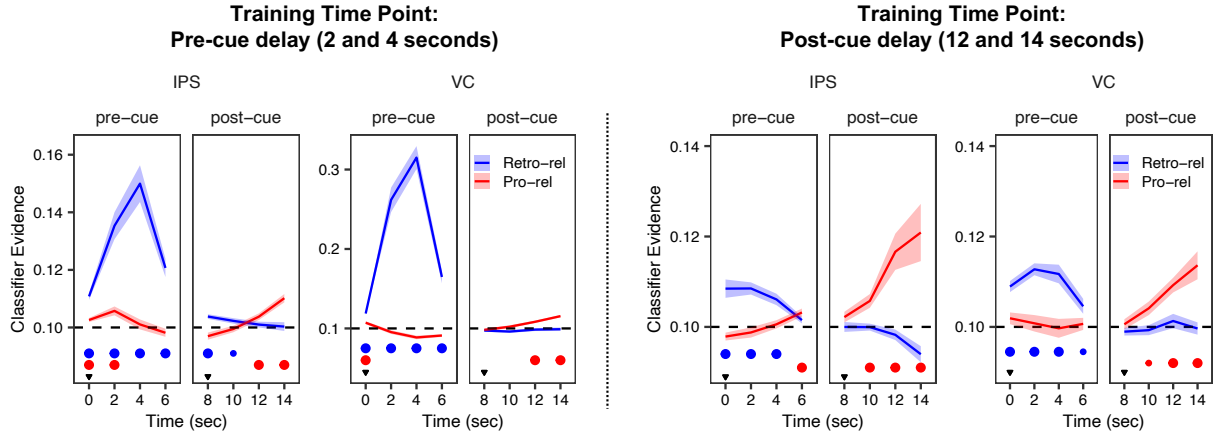

Fig 3. Tracking retrospective and prospective codes over time. Separate classifiers were trained using data from the pre-cue (2 and 4 seconds) and post-cue (12 and 14 seconds) delays (see red dotted boxes in **Fig 2b**). Retrospective-relevant ('Retro-rel') refers to the classifier evidence corresponding to the spatial location of the just-probed item. Prospective-relevant ('Pro-rel') refers to the classifier evidence corresponding to the spatial location following the reference (see **Fig 1c**). The significance at each time point is denoted by small and medium dots, representing q < 0.05 and q < 0.01, respectively (FDR corrected). The inverted triangles denote the time points of the presentation of the probe and the cue, respectively.

To more directly detail the nature of WM transformations, we examined the distribution of classification evidence across time (pro- vs. retrospective) and priority (relevant vs. irrelevant hemifield). Specifically, we separately detailed the classifier evidence for the retrospective spatial location in the relevant hemifield (retrospective-relevant; 'Retro-rel'), the retrospective spatial location in the irrelevant hemifield (retrospective-irrelevant; 'Retro-irrel'), the prospective spatial location in the relevant hemifield (prospective-relevant; 'Pro-rel'), and the prospective spatial location in the irrelevant hemifield (prospective-irrelevant; 'Pro-irrel') (see **Fig 1c** for this categorization). This allowed us to

determine the extent to which representations transformed over time and whether there was representational evidence of the non-prioritized hemifield. For the pre-cue period, the relevant hemifield was considered the just-probed hemifield, while in the post-cue period, the cued hemifield was considered the relevant hemifield. Anticipating the possibility that pre- and post-cue representations may differ [44], we trained separately classifiers using the peak classification time points of the pre-cue delay (2 and 4 seconds) and the peak time points of the post-cue delay (12 and 14 seconds) and tested each across all time points, respectively (red box outlines in **Fig 2b**). In this case, the pre-cue trained representations contain a mixture of sensory (from the visual presentation of the probe) and WM signals, while the post-cue trained representations should reflect purely WM signals. If representations supporting WM mainly reflect sensory information, similar results would be expected based on each training period – but if WM representations also contain non-sensory information, the results across delay period should differ [17].

Figure 3 depicts the classifier evidence for the retrospective-relevant and prospective-relevant locations averaged across the repeat and the switch conditions. In both the IPS and VC, the classifier evidence for the retrospective-relevant spatial location was significantly above chance levels during the pre-cue delay, whereas it dropped to chance level during the post-cue delay. By contrast, the classifier evidence for the prospective-relevant spatial location showed the opposite pattern suggesting a shift from a retrospective to prospective code. (Note that there was some transient evidence for the prospective-relevant spatial location during the pre-cue period, but this may be attributable to non-specific attention to the relevant hemifield which we will account for in analyses described below.) Most evidence for either the retrospective-irrelevant or prospective-irrelevant spatial locations were under the chance level (**Fig S2**), indicating that decodable representations in the IPS and VC were restricted to prioritized items as consistent with substantial past work [20, 43-46]. In subsequent analyses, we will therefore focus on relevant items.

Taken together, the results indicate that WM was transformed over the delay periods from retrospectively representing the past to prospectively representing expectations for the future.

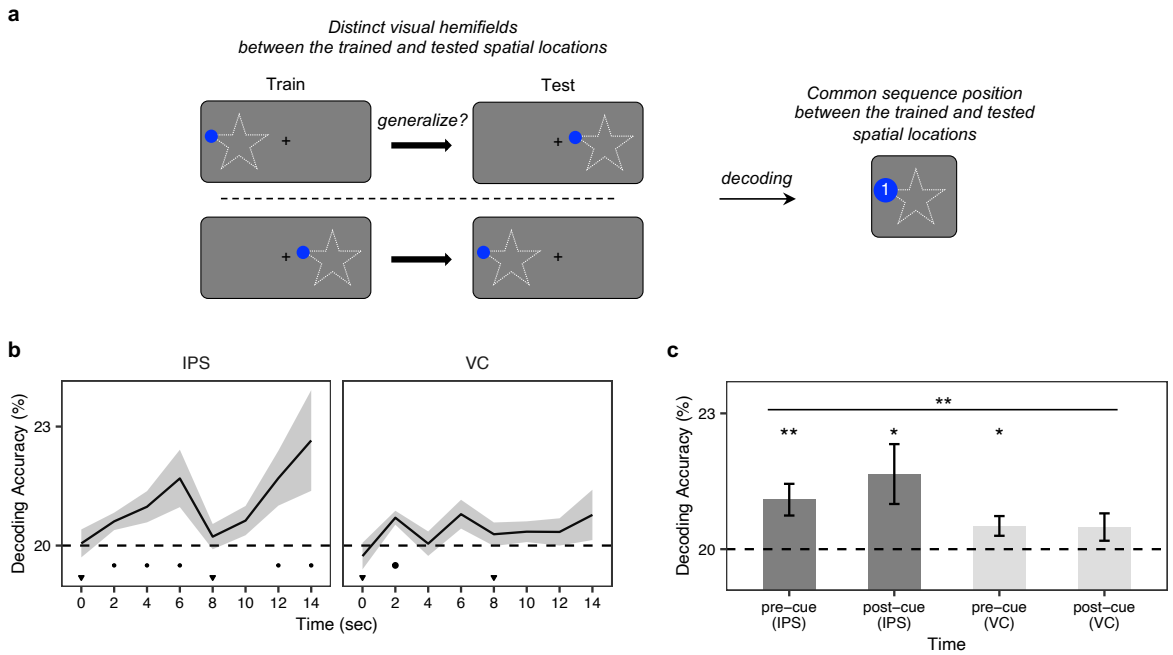

Fig 4. Decoding abstract sequence positions.
(a) Abstract coding was operationally defined based on generalization of decoding across visual hemifields.
(b) The result of the cross-visual hemifield decoding analysis, in which training and testing the classifier were performed on different visual hemifields, respectively. Decoding performance was compared against the chance level of 20% and corrected for multiple comparisons using false discovery rate (FDR). The significance at each time point is denoted by small, and medium dots, representing $q < 0.05$, and $q < 0.01$, respectively. The shaded regions represent ± 1 standard error of mean. The inverted triangles in black denote the time points of the presentation of the probe stimulus and the cue, respectively.
(c) Average of decoding accuracies from 'pre-cue' (2 sec to 6 sec) and 'post-cue' (10 sec to 14 sec) period in each ROI. The horizontal dashed line indicates the chance level of 20%. Each dot on the bar graph represents a participant. $^*p < 0.05$, $^{**}p < 0.01$. Error bars represent ± 1 standard error of mean.

Abstract sequence information is coded in the IPS.

In the current task, expectations in each visual hemifield were guided by the same abstract spatial sequence. Abstraction is commonly defined as generalization across contexts [27, 28, 37]. Here, each visual hemifield formed a distinct context. To test for abstract coding of sequence, we examined the extent to which representations generalized across hemifields by training a classifier using representations from one hemifield and testing it with representations from the other hemifield (cross-visual hemifield decoding; **Fig 4a**). Training and testing were performed separately for each TR

using labels of the relevant sequence position. As shown in the Figure 4b, we found significant cross-visual hemifield decoding during both the pre-cue and post-cue delays in the IPS with decoding accuracy ramping over time during both phases (**Fig 4b**). By contrast, in the VC, the classifier showed above-chance decoding accuracy only during a single time point in the pre-cue delay. To better quantify this result, we averaged the decoding accuracies from 2 sec to 6 sec to form the 'pre-cue' period and from 10 sec to 14 sec to form the 'post-cue' period separately for each region. These data were submitted to a period ('pre-cue' vs. 'post-cue') by region (IPS vs. VC) RM ANOVA. This analysis revealed a significant main effect of region with superior decoding accuracy in the IPS relative to the VC ($F_{(1, 14)} = 8.74$, $p = .008$) (**Fig 4c**). Neither the main effect of period nor the interaction between period and region was significant. Follow-up t-tests revealed significant decoding in both the 'pre-cue' (21.1%, $p = .004$) and 'post-cue' periods (21.7%, $p = .014$) in the IPS. By contrast, in the VC, decoding accuracy was different from chance in the 'pre-cue' (20.5%, $p = .017$), but not in the 'post-cue' (20.5%, $p = .064$). Collectively, although there was some weak evidence for abstract codes in the VC, abstract codes were prominent in the IPS.

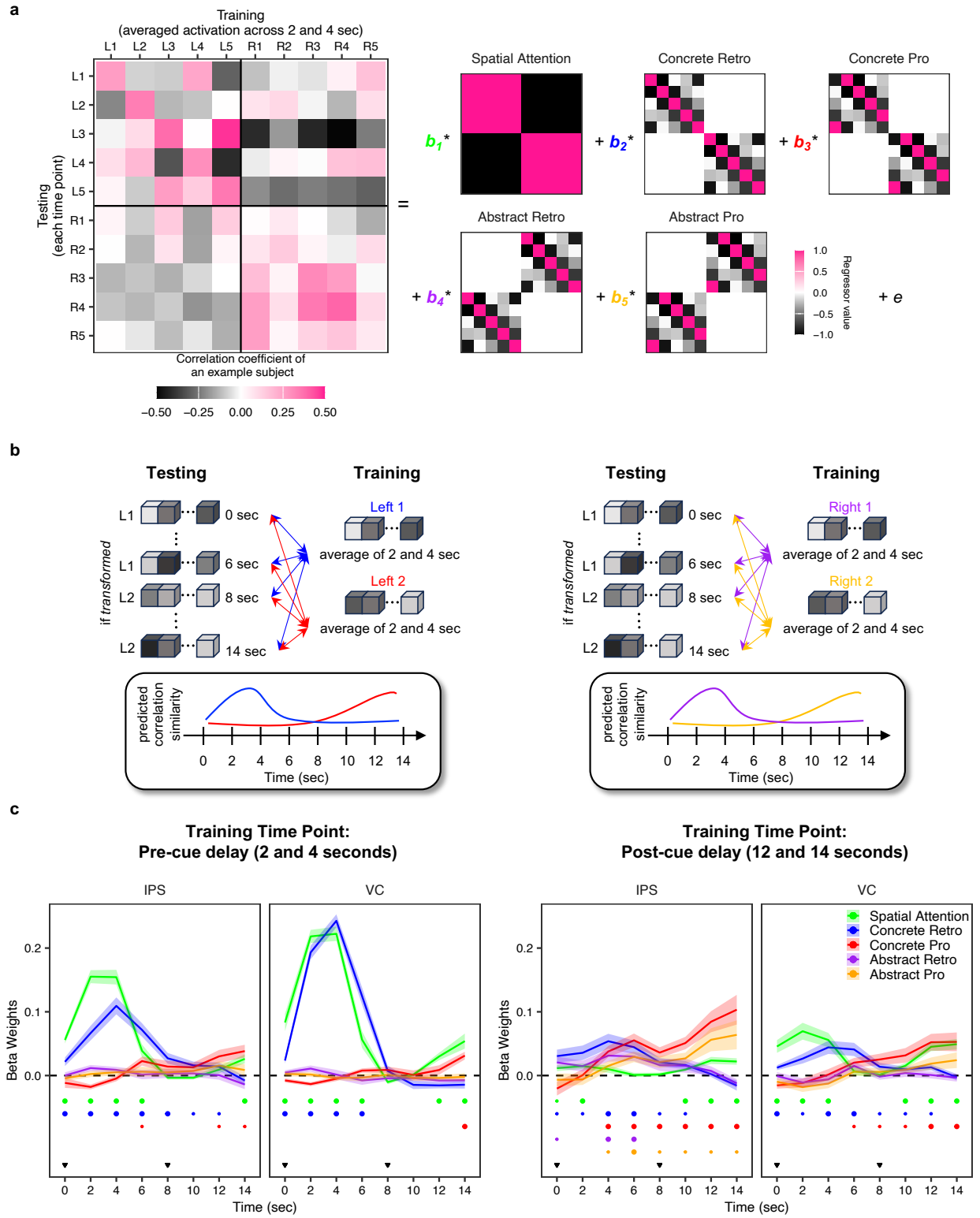

Fig 5. Concrete and abstract codes of working memory are transformed over time.

(a) The left depicts an example representational similarity matrix formed by correlating training patterns of each location with testing patterns of each location. These matrices

were submitted to multiple linear regression using regressors denoting idealized matrices of spatial attention, concrete retrospective coding, concrete prospective coding, abstract retrospective coding, and abstract prospective coding. We estimated the beta weight for each regressor at each time point.
(b) Schematic of the predicted correlation similarities representing a transition between concrete retrospective to concrete prospective coding (left) and abstract retrospective to abstract prospective coding (right). (See also **Fig S5** for the analogous regressors for the post-cue delay trained data).
(c) The estimated beta weights for each regressor at each time point. The beta weights were compared against a value of 0 and corrected for multiple comparisons using false discovery rate. The significance at each time point is denoted by small and medium dots, representing $q < 0.05$ and $q < 0.01$, respectively (FDR corrected). The shaded regions represent ± 1 standard error of mean. The inverted triangles in black denote the time points of the presentation of the probe stimulus and the cue, respectively.

Simultaneous and dynamic coding of the past and future.

The analyses thus far have demonstrated the presence of concrete and abstract codes underlying WM. To better examine the extent to which these codes are multiplexed versus trade off over time, we turned to representational similarity analysis (RSA). For each of the 10 spatial locations, we extracted the patterns of activation in each of our 4 ROIs (left and right IPS, left and right VC). To form training patterns, we separately averaged over the 2 and 4 second post-stimulus TRs ('pre-cue') and the 12 and 14 second post-cue TRs ('post-cue') as we did above. Then in a testing set, for each TR, we calculated the correlation similarity between the activity pattern at that TR and each training pattern (**Figs 5a and S3a**). This was done separately for each ROI. We then decomposed representations in the IPS and VC into 1) coarse representations of the attended hemifield (spatial attention), 2) fine-grained representations of the retrospective, relevant spatial location (concrete retrospective), 3) fine-grained representations of the prospective, relevant spatial location (concrete prospective), 4) abstract representations of the retrospective sequence position (abstract retrospective), and 5) abstract representations of the prospective sequence position (abstract prospective) using multiple linear regression. Here, each regressor represents a representational similarity matrix corresponding to a hypothetical pure representation (see Methods; **Figs 5a and S3a**). The regressors incorporated distance-dependent similarity based upon normalized visual angle (i.e., nearby locations are more similar to one another than far locations). Concrete

regressors were formed by training/testing in the same hemifield, while abstract regressors were formed by training/testing across hemifields. Based on the analyses above, we expected that retrospective codes would transition to prospective codes both concretely and abstractly (**Fig 5b; Fig S3b**).

As shown in Figure 5c, we found evidence consistent with these predictions. In both the IPS and VC, a concrete retrospective code emerged early during the 'pre-cue' delay, but diminished over time until it was absent by the end of the 'post-cue' delay. Concrete retrospective codes were most prominently observed in the 'pre-cue' trained data. The fall of the concrete retrospective code was mirrored by a rise of a concrete prospective code which started before the cue [47] and peaked at the end of the 'post-cue' delay. Prospective coding was more prominently observed in the 'post-cue' trained data. Moreover, we found evidence for an abstract retrospective code during the 'pre-cue' delay which was clearest in the IPS using the 'post-cue' trained data. Interestingly, abstract retrospective codes in the IPS rose in time with abstract prospective codes during the 'pre-cue' delay (i.e., 4 and 6 s) suggesting that some transformations occur prior to the cue . Abstract prospective codes remained sustained in the IPS throughout the 'post-cue' delay while concrete prospective codes were sustained in both the IPS and VC during the 'post-cue' delay.

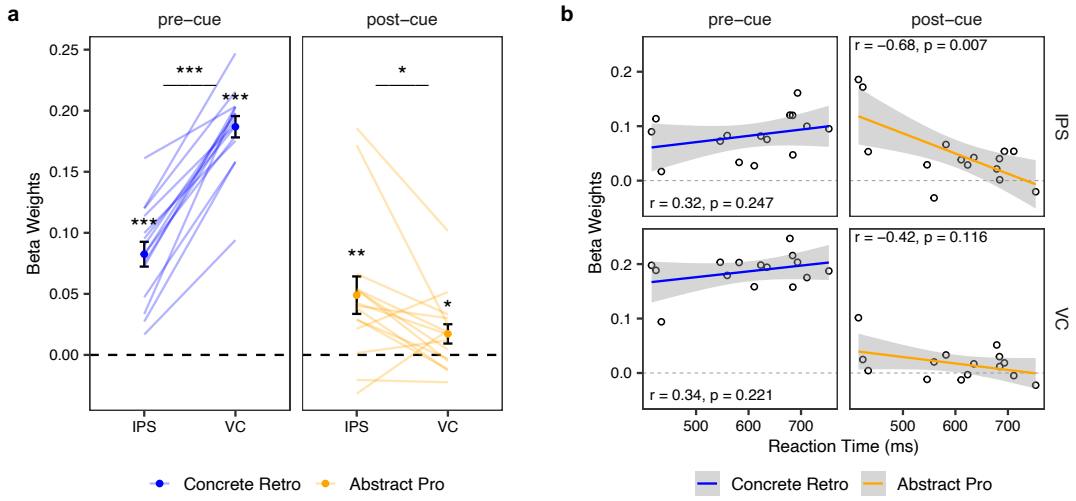

Fig 6. Different roles of IPS and VC in representing WM codes.

(a) Contrasting concrete retrospective (Concrete Retro) and abstract prospective (Abstract Pro) codes in the intra-parietal sulcus (IPS) and visual cortex (VC). The concrete retrospective code was formed by averaging beta weights during the 'pre-cue' period (2 sec to 6 sec) for each ROI using the pre-cue trained data (blue). The abstract prospective code was formed by averaging beta weights during the 'post-cue' period (10 sec to 14 sec) for each ROI using the post-cue training period (orange). Each line represents each participant. Bold color dots indicate the mean across participants with error bars representing ± 1 standard error of mean. *p < 0.05, **p < 0.01, ***p = 0.001.
(b) Correlations between the strengths of concrete retrospective (blue) and abstract prospective (orange) and RT for each ROI. Each dot represents each participant. The shaded regions represent 95% confidence interval of the correlation.

The results above suggest that concrete retrospective coding was more dominant in the VC than in the IPS, while abstract prospective coding was more dominant in the IPS than in the VC. To directly quantify the different roles of IPS and VC in representing WM codes, we averaged the beta weights of the concrete retrospective code from the 'pre-cue' (2 sec to 6 sec) period using the pre-cue trained data, and the beta weights of the abstract prospective code from 'post-cue' (10 sec to 14 sec) using the post-cue trained data (**Fig 6a**). Then, we ran a two-way RM ANOVA with the representation ('concrete retro' vs. 'abstract pro') and region ('IPS' vs. 'VC') as within-subject factors. We observed a main effect of representation such that concrete retrospective codes were more strongly detected than abstract prospective codes ($F_{(1, 14)} = 68.94$, $p = .001$). We also observed a main effect of region such that representations were generally more strongly observed in VC than IPS ($F_{(1, 14)} = 20.80$, $p = .002$). Critically, we observed a significant interaction between the two factors ($F_{(1, 14)} = 108.15$, $p = .001$). Although VC showed stronger concrete retrospective coding than the IPS ($t_{(14)} = -13.21$, $p = .001$), the reverse was true regarding abstract prospective coding ($t_{(14)} = 2.59$, $p = .025$). These data indicate a dissociation with VC more strongly coding for the concrete past and the IPS more strongly coding for the abstract future.

If abstract coding in the IPS is useful for guiding future behavior, stronger abstract coding should be related to better behavioral preparedness (i.e., faster RTs). To test this idea, we examined the extent to which concrete retrospective and abstract prospective codes related to RT (**Fig 6b**). While concrete retrospective coding was non-significantly positively related to RT in both regions (r's < 0.4, *ps* > .22), abstract prospective coding

was negatively related to RT with a significant relationship observed in the IPS (r = -0.68, *p* = .007), but not VC (r = -0.42, *p* = .116). This pattern was observed in both 'Repeat' and 'Switch' trials suggesting that stronger abstract coding was related to general preparedness (**Fig S4**). Relationships to ER trended non-significantly in the same direction indicating that the pattern could not be attributed to a speed-accuracy tradeoff. Additionally, although both concrete and abstract prospective codes in the IPS were concurrently maintained during the post-cue delay in the 'post-cue' trained data (**Fig 5c**), there was no significant relationship between the strength of concrete prospective coding and behavior (**Fig S4c**). Collectively, these data suggest that abstract coding in the IPS facilitates future behavioral performance.

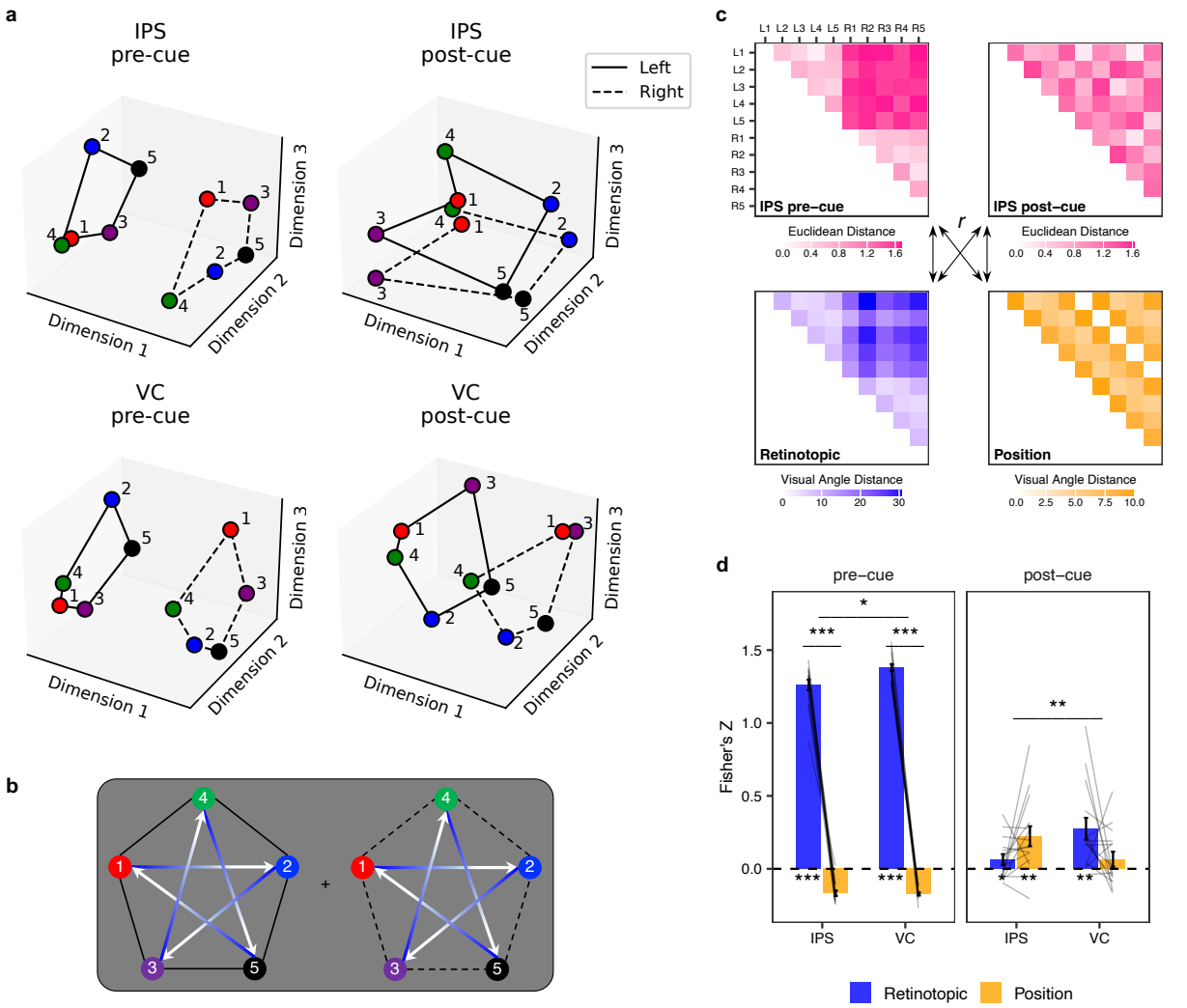

Fig 7. Representational geometry of WM.
(a) Each of the 10 spatial locations is plotted as a function of the first three dimensions determined by multidimensional scaling (MDS). Spatial locations are separately color coded by sequence position. Spatial locations in the left and right visual hemifields are connected by solid and dotted black lines, respectively. (b) Schematic illustration of the spatial locations. (c) Group-level Euclidean distance in the multidimensional space between spatial locations in the IPS pre- and post-cue (top row). Visual angle distance between spatial locations (bottom row). The visual angle distance was estimated considering retinotopic (bottom left) and sequence position geometries (bottom right). (d) The result of correlation between Euclidean distance and visual angle distance. Each line on the bar graph represents each participant. *p < 0.05, **p < 0.01, ***p = 0.001. Error bars represent ±1 standard error of mean.

WM representational space is reorganized over time.

The RSA analyses revealed multiplexing of representational codes supporting WM in both the IPS and VC. To more directly visualize these representations, we turned to multidimensional scaling (MDS) which helps to elucidate representational geometry. For each delay period and region, we submitted the similarity structure of the 10 spatial locations to MDS and plotted each location in the space formed by the first three dimensions (see Methods). This revealed a clear geometry during each period, and transformation of the representations supporting WM over time. During the 'pre-cue' period, representations were separated by hemifield along the first dimension in both the IPS and VC (**Figs 7a and S6b**). For both areas, a clear circular geometry was present across the second and third dimensions resembling the structure of the visually-presented locations (**Figs 7b and S6c**). Across hemifields these circular structures were vertically aligned but mirror-reversed consistent with distance in retinotopic space (e.g., position 2 in the left hemifield is close to position 1 in the right hemifield; **Fig 7b**). Hence, retinotopic space was well-recapitulated across the three dimensions of MDS in both the IPS and VC.

While a similar representational geometry was observed in the IPS and VC during the 'pre-cue' period, this was not the case during the 'post-cue' period. In the 'post-cue' period, hemifield discrimination was reduced and relegated to the third dimension in the IPS. A circular geometry resembling the star-sequence was observed along the first and second dimensions (**Fig S6b**). Importantly, in contrast to the retinotopic organization observed during the 'pre-cue' period, these circular geometries were aligned by position across hemifields. In VC, hemifield discrimination remained prominent along the first dimension. Alignment by sequence position was also observed in VC, albeit across the second and third dimensions. Hence, representational transformations were observed in both the IPS and VC transitioning from a concrete, retinotopic code to an abstract, position code. However, the areas differed in their principal coding axis. The IPS more prominently coded for sequence position while the VC more prominently coded for hemifield.

To quantitatively evaluate the representational dynamics of the MDS, we examined the extent to which Euclidean distance among representations in the MDS space resembled

retinotopic distance and abstract position distance. Retinotopic distance was calculated as the distance in visual angle among spatial locations. Position distance was calculated similarly, but assumed that positions across hemifields were identical (**Fig 7c**). For each subject, we correlated the matrix of Euclidean distances in MDS space with the matrix of Euclidean distances in retinotopic and position space, respectively, separately for the pre- and post-cue delays. Correlation values were z-transformed and submitted to a period (‘pre-cue’ vs. ‘post-cue’) by region (IPS vs. VC) by space (retinotopic vs. position) RM ANOVA. This analysis revealed a significant three-way interaction ($F_{(1, 14)} = 6.41$, $p < .05$; see Table S1 for complete stats). Specifically, in the ‘pre-cue’ period, WM representational geometry closely resembled retinotopic space much more so than position space in both the IPS and VC (‘pre-cue’ period main effect of space; $F_{(1, 14)} = 2344.48$, $p < .001$). By contrast, in the ‘post-cue’ period, the IPS representational geometry was more similar to the sequence position space than the retinotopic space, whereas the reverse was true in the VC representational geometry (‘post-cue’ period region x space interaction: $F_{(1, 14)} = 17.65$, $p = .002$). Taken together, these results suggest that WM representational geometries were reorganized over time. Specifically, both the IPS and VC were dominated by a concrete, retinotopic code in the ‘pre-cue’ delay. By contrast, in the ‘post-cue’ delay, the IPS primarily coded for abstract sequence position, whereas the VC maintained a more concrete retinotopic-dependent code.

Discussion

Although the primary function of WM is to hold past sensory information to guide future cognition [8], the mechanisms supporting the transformation of past sensory codes into abstract codes that guide future cognition have been elusive. Here, we found evidence for concrete, retrospective codes in the VC and the IPS when human participants maintained past sensory information (pre-cue delay). However, when anticipating a forthcoming probe (post-cue delay), this concrete, retrospective code disappeared concomitant with the appearance of a prospective code. Using cross-visual hemifield decoding, we found generalization of spatial position coding across visual hemifields which ramped over time in the IPS consistent with a rise in abstract coding to inform transformations from the last position to the next. RSA further detailed this finding, showing that retrospective codes transition to prospective codes. Moreover, the strength of abstract prospective coding in the IPS was related to more efficient behavioral performance consistent with a signal that prepares for future cognition. Lastly, dimensionality reduction (MDS) revealed how the representational geometry for WM in both regions changed over time. Specifically, WM codes re-organized from retinotopic to being aligned by abstract sequence position. Representational geometries in the VC and IPS were highly similar prior to the cue, but diverged substantially following cue-driven transformations with the IPS emphasizing abstract, sequence positions to guide future responses. Taken together, these results detail how WM transforms to support future cognition with dissociable roles of VC more strongly representing the concrete past and the IPS more strongly representing the abstract future.

Numerous previous studies have revealed that concrete retrospective sensory signals can be decoded in the IPS and VC during WM delay periods [10, 13, 14, 17, 18, 20]. These findings have led to the conclusion that both regions serve as critical sites for the storage of visual WM. However, a central function of WM is the transformation of past information into future-relevant information beyond just accurately reproducing or recognizing past information. Recent studies using electroencephalography or eye-tracking have provided evidence showing that anticipated actions [29, 32] and spatial locations [48, 49] are prospectively coded during WM delays. Furthermore, human fMRI

503 studies have decoded prospective category-level WM representations in areas
504 processing visual objects such as fusiform gyrus, middle temporal gyrus, and lateral
505 occipital cortex [46, 50]. Building on these findings, our data show that the fine-grained
506 representation of past spatial locations in the IPS and VC are transformed into anticipated
507 spatial positions when preparing for a sequential probe. Moreover, we show that although
508 expectations of the future have a spatial geometry that resembles some aspects of
509 sensation, this geometry is abstracted away from sensation such that locations that are
510 retinotopically distant can nevertheless align in representation space when they share
511 meaningful features. This abstraction may explain why training classifiers during
512 perception can lead to significant decoding of WM, but that training classifiers during WM
513 results in superior decoding of WM [14, 17, 44]. This result implies that the IPS and VC
514 store not only the concrete past but also the anticipated future with the future stored in a
515 more abstract format than the past.
516
517 Although several works have shown similar contributions of the IPS and VC to visual WM,
518 other work has suggested that their roles may be dissociable. For example, some studies
519 have demonstrated that the presence of a distractor during a WM delay disrupts or biases
520 WM representations in the VC, whereas WM representations in the IPS are robust against
521 distractors [10, 16, 17]. This work implies that the IPS may store not only concrete,
522 sensory-like WM codes but also non-stimulus-driven WM representations that are
523 transformed away from the sensory signals [17]. Consistent with this idea, our data show
524 evidence of both concrete sensory and abstract position signals in the IPS with concrete
525 representations transitioning into abstract representations over time. By contrast,
526 evidence for such transitions in the VC were weaker or non-significant depending upon
527 the exact form of the analysis. We suggest that the hints of such signaling in the VC may
528 be due to top-down signals from the IPS. Such an idea is consistent with recent work
529 showing that during WM delays, the IPS shapes the contents of VC through feedback
530 [51]. Moreover, not only were abstract prospective codes consistently stronger in the IPS
531 than VC, the strength of abstract prospective codes in the IPS were related to improved
532 behavior. Collectively, these data indicate dissociable roles of the VC and IPS with the

former involved in concrete, sensory representation and the latter important when abstract, prospective codes facilitate performance.

Recent work has shown that VC representations can also be abstracted away from a strict sensory format. VC representations supporting WM can generalize across gratings and motion directions into a line-like format [28], and can show invariance to aperture biases that affect sensory representations [27]. In those works, generalization was examined within the same spatial receptive field. Here, we examined generalization across spatial receptive fields (i.e., across visual hemifields). Hence, that abstract coding was weak to absent in the VC in our data does not undermine the presence of other forms of abstract coding in the VC. Rather, there may well be different levels of abstraction which transition across the visual hierarchy. Since we did not perform retinotopic mapping to carefully tease apart different levels of the visual hierarchy here examination of how abstract coding in the present task varies along the levels of the visual hierarchy is a question for future work.

According to a recent proposal, the nature of abstraction in WM is to transform its representational structure—reflecting past informational content—into a future-oriented format to guide the upcoming task [52-55]. Panichello and Buschman [40] contrasted selection from WM with selection from perception (i.e., attention) in monkeys. They found that selection in both forms generalized in the prefrontal cortex (PFC), but not parietal cortex or VC, providing evidence for an abstract mechanism of control in the PFC. Moreover, PFC representations of WM items transformed from a pre-cue subspace wherein WM representations for different colors were multiplexed with location, to a post-cue subspace wherein representations for different colors aligned regardless of location. The same transformation in representational geometry was also observed in a recurrent neural network (RNN) trained to carry out an equivalent cued-recall paradigm [41, see also 42]. Here, we observed a similar transformational alignment, particularly in the IPS. Observing these signals in the IPS here may be due to the fact that spatial information remained relevant for the task and that the IPS and other dorsal stream areas are important for the representation of space [56]. These findings align with the proposition

that abstraction in WM is reflected in the transformation of representational structure into a future-oriented format.

In sum, a combination of multivariate decoding, RSA, and MDS revealed that both IPS and VC initially maintained concrete past locations with concrete, retrospective coding particularly prominent in the VC. Over time, these representations became progressively more abstract and prospective with abstract, prospective coding particularly prominent in the IPS. These results suggest that WM guides future cognition by reformatting sensory signals of the past into more abstract codes to guide prospective expectations.

Acknowledgments

The authors would like to thank Michael Dicalogero for help with collecting the data. This work was supported by startup funds from Florida State University (DEN).

Author contributions

DD and DEN designed the experiment. DD collected the data. DD and JL analyzed the data with input from JAM and DEN. JL and DEN wrote the manuscript with input from DD and JAM.

Declaration of interests

The authors declare no competing interests.

Methods

Participants

22 participants were recruited from the Tallahassee area for this experiment. All participants underwent an initial screening process to ensure that they understood and could adequately perform the task. Informed consent was obtained according to the guidelines set forth by the Office for Human Subjects Protection and Florida State University Institutional Review Board. For the experiment, a total of three scanning sessions were planned. 7 participants were removed from the study following the first scanning session due to insufficient data quality (e.g., poor task performance, excessive head motion, poor univariate activation). These data were excluded from further analyses. The remaining 15 participants (age range 18-22; mean age: 19,8; 10 female) were analyzed for this report. 1 participant did not complete the last scanning session due to a technical issue with the remaining participants completing all three sessions.

fMRI experiment procedure

In the fMRI scanner, participants performed a sequence working memory task. In the task, participants were required to maintain two spatial locations, one in each visual hemifield, each of which were drawn from a five-location sequence tracing a star (**Fig 1**). These stimuli were used as references for the probes. On each trial, a cue was presented at the center of the screen for 500 ms, pointing to either the left or right visual hemifield, indicating that an upcoming probe would appear in that respective visual hemifield (task-relevant visual hemifield). For all but the first cue, cues were preceded (pre-cue delay) and followed (post-cue delay) by 7500 ms of fixation. After the post-cue delay, a probe was presented in the task-relevant visual hemifield for 500 ms. Participants were asked to determine whether the spatial location of the currently presented probe sequentially followed the reference (i.e., the spatial location of the last stimulus presented in the same visual hemifield). The spatial location of the probe then was used as the reference for the next probe in the same visual hemifield. Thus, the reference in the probed hemifield was updated in each trial.

Half of the probes followed the reference in the sequence thereby requiring a “match” response. “Non-match” probes were pseudo-randomly and equally drawn from the three spatial locations remaining after excluding the sequence “match” and the reference (i.e., there were no stimulus repeats). Trials were equally distributed to cue the left and right visual hemifields and to require even numbers of switches and repeats of the visual hemifield of the preceding trial. Each run began with the presentation of two location stimuli for 1500 ms which served as the initial reference locations. A delay of 2500 ms separated the initial stimuli from the first cue. During the experiment, participants were asked to fixate their gaze on a fixation cross the center of the screen and not move their eyes. Adherence to these instructions was monitored by eye tracking. 14 out of 15 participants completed three scanning sessions, while one participant completed only two scanning sessions due to a technical issue. Each scanning session was carried out on separate days. Each scanning session consisted of 4 experimental runs and 50 trials per experimental run. There was a short break between experimental runs.

Stimuli and apparatus

The experiment was programmed with E-Prime software version 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA). The probe stimulus was a blue circle with a diameter of ~2°. Stimuli were presented between 5.5° and 15.25° of visual angle left and right and between 3.32° and 4.68° of visual angle above and below a central fixation point.

MRI acquisition

MRI data were collected on a Siemens 3T Prisma with a 32-channel head coil. The participant observed visual stimuli in the scanner using a mirror attached to the head coil, which reflected the visual stimuli onto a projection screen. Responses were recorded using a 4-button MR-compatible button box (Current Designs, Inc.), with participants using the index finger of each hand to input their responses.

Functional imaging data were obtained using an EPI sequence provided by the Center for Magnetic Resonance Research (CMRR) at the University of Minnesota: voxel dimensions of 2 × 2 × 2 mm, TR = 2000 ms, TE = 33.8 ms, flip angle = 45°, FOV = 192

645 mm^2, multi-band factor = 4. Prior to each functional scan, four dummy scans were
646 conducted to ensure image stabilization. Phase and magnitude images at the same
647 resolution of the functional images were acquired to assess and correct for magnetic field
648 inhomogeneity. Additionally, a high-resolution T1-weighted MPRAGE image was
649 collected for spatial normalization (384 × 384 ×256 matrix of 0.667 mm3 isotropic voxels;
650 TR =1840 ms; echo time = 2.9 ms; flip angle = 9°).
651
652 **MRI data preprocessing**
653 Preprocessing on image data was done in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/)
654 unless otherwise specified. DICOM format was converted into NIfTI format. The origin of
655 all images in each participant was manually adjusted to the anterior commissure.
656 Functional image data underwent spike correction to mitigate the influence of artifacts
657 using AFNI's 3dDespike routine (http://afni.nimh.nih.gov/afni), as well as slice timing
658 correction for each run, head motion correction via a six-parameter rigid body
659 transformation, and unwarping and correction of motion-by-susceptibility distortions using
660 the FieldMap toolbox [57]. Based on the realignment parameters, linear, squared,
661 differential, and squared differential movement parameters were calculated [24 in total;
662 58]. These parameters were regressed out of the functional images and the resultant
663 residuals were high-pass filtered at 1/128 Hz. These data were used for all described
664 analyses.
665
666 Structural image data were coregistered to the functional image data. Then, segmentation
667 was performed on structural image data to acquire gray and white-matter probability maps
668 [59] from which spatial normalization warping to and from the MNI template was
669 calculated.
670
671 **ROI selection**
672 We selected the left and right intra-parietal cortex (IPS) along with the left and right visual
673 cortex (VC) as the regions of interest (ROIs) given their role in maintaining working
674 memory [10, 13, 14, 16-18]. ROI masks were created using a probabilistic atlas [60] with
675 a 90% probability threshold. VC ROIs were created by combining V1v, V1d, V2v, V2d,

V3v, V3d, and hV4. IPS ROIs were created by combining IPS0-5. These ROIs were warped to native space for each individual.

Multivariate fMRI analyses

The BrainIAK toolbox implemented in Python was used to conduct decoding analyses [61]. All decoding analyses were performed on preprocessed BOLD responses from the ROIs of each participant. To appropriately balance factors for classification, consecutive pairs of runs were concatenated. The concatenated BOLD data of each ROI were temporally z-scored to ensure that voxel activities were scaled to the same range. Only correct trials were analyzed. After removing error trials, we equalized the number of trials of each spatial location in each run to match the smallest count among them by randomly sampling the larger number of trials of each spatial location resulting in ~8 trials per concatenated run per location on average. Such random sampling and cross-validation were repeated 1000 times.

To train and test a classifier (L2-regularized, multinomial logistic regression), we implemented a leave-one-pair-out cross-validation procedure. The decoding accuracy of each participant was calculated as the average decoding accuracy across all iterations. During cross-validation iteration, we applied univariate feature selection using an ANOVA (f-classif function in the scikit-learn Python library [62] with a threshold alpha level of 0.05 on each training data. The features selected based on the training data were used on the testing data. Since this procedure resulted in a variable number of features as a function of ROI and time point, we repeated the analyses using a fixed number of features (300, which approximated the lower bound of the number of features selected by the above procedure) which produced the same patterns of results as reported in the main text (**Figs S5a, b, and c**) indicating that the results are robust to the specifics of feature selection. The L2 penalty was the value of 1, which was used to train the classifier. For the decoding analysis, we excluded the initial trial of each run as the timing and visual presentation of this trial differed from the others.

Multivariate fMRI analyses: temporal generalization decoding analysis

Temporal generalization decoding analysis [39] was conducted to assess whether the representation of the reference remained stable or transformed over time. To examine this question, a classifier was trained to classify 10 distinct categories representing the spatial locations of stimuli (five locations in each hemifield) at each TR (8 TRs in total; from 0 to 14 s from the onset of the probe). Then, each trained classifier was tested to classify each TR of the independent testing data (8 TRs in total; from 0 to 14 s from the onset of the last probe). For repeat trials, we assigned the same category label to 8 TRs of the pre-cue and post-cue delays, and the category of the label corresponded to the probe of the preceding trial. For switch trials, the 4 TRs of the pre-cue delay were labeled in the same way. However, the 4 TRs of the post-cue delay were labeled with the last probe presented on the cued visual hemifield. In other words, the pre-cue delay was always labeled with the last probe while the post-cue delay was always labeled with the reference for the upcoming probe.

Training and testing were performed separately for each of the four ROIs (left and right IPS and VC). Given similarity in the results in each hemifield, left and right hemifields were averaged together to summarize a given brain area (e.g., left and right IPS results were averaged to characterize the IPS).

Multivariate fMRI analyses: tracking retrospective and prospective codes over time

The temporal generalization decoding analysis produced classifier evidence corresponding to each of the 10 spatial locations at each time point. We specifically tracked evidence for the retrospective-relevant (‘Retro-rel’), retrospective-irrelevant (‘Retro-irrel’), prospective-relevant (‘Pro-rel’), and prospective-irrelevant (‘Pro-irrel’) locations (**Fig 1c**). During the pre-cue delay, the just-probed spatial location was categorized as ‘Retro-rel’, and the last probed spatial location in the uncued hemifield was categorized as ‘Retro-irrel’. During the post-cue delay of the repeat condition, the next spatial locations of ‘Retro-rel’ and ‘Retro-irrel’ in the sequence were categorized as ‘Pro-rel’ and ‘Pro-irrel’, respectively. In the switch condition, the classifier evidence corresponding to the next spatial locations of ‘Retro-irrel’ and ‘Retro-rel’ in the sequence were categorized as ‘Pro-rel’ and ‘Pro-irrel’, respectively. For each time point, classifier

evidence for each of the four categories of interest were averaged across trials for each subject and submitted to statistical analysis.

Multivariate fMRI analyses: cross-visual hemifield decoding analysis

Cross-visual hemifield decoding analysis was performed by training a classifier on the five sequence positions presented on the one visual hemifield (e.g., the left (right) visual hemifield) and testing the classifier on the five sequence positions presented on the other visual hemifield (e.g., the right (left) visual hemifield). Decoding was conducted in a time-resolved manner, involving training and testing the classifier within the same TR. As we did in the temporal generalization analysis, the label at each TR during the pre-cue delay was based on the just-probed stimulus, while the label at each TR during the post-cue delay was based on the last probed stimulus in the cued hemifield.

Representational similarity analysis

We conducted a representational similarity analysis (RSA) [63] on the preprocessed BOLD data to examine the transformation of the neural representation of WM across time points. To setup the data for RSA, consecutive pairs of runs were concatenated in the same way as in the decoding analyses. Then, data were split into training and testing sets by using a leave-one-pair-out-cross-validation procedure. For each iteration of the cross-validation, univariate feature selection was performed using an ANOVA with a threshold alpha level of 0.05 on the training data (note similar patterns of results were obtained by fixing the number of features at 300, **Fig S5d**). The selected features (voxels) were then z-scored as a collection, followed by z-scoring within each feature. Based on these normalized features on the training data, we selected samples corresponding to the 10 spatial locations at the peak decoding time points of the pre-cue delay (2 and 4 seconds; the second and third TR after the presentation of the last probe) and the peak decoding time points of the post-cue delay (12 and 14 seconds; the third and fourth TR after the presentation of the cue). Then, for each period, we calculated the average value of the selected samples for each spatial location to form the training representations (10 spatial locations x 2 periods). For the testing data, we used the same features selected on the training data and applied z-score normalization in the same manner as was done for the

training data. Then, we selected samples corresponding to the 10 spatial locations for each time point. For each time point, we averaged the patterns of activation for each of the 10 spatial locations, respectively, to form the testing representations. Then, for each time point, we then calculated the correlation coefficient among the 10 training representations and the 10 testing representations, separately for each training period set. In other words, for each training period, eight RSA matrices were formed, one per time point, showing the representational similarity between the 10 spatial locations of the training set and the 10 spatial locations in the testing set. Each RSA matrix was submitted to model-based regression analysis (see below) for each iteration of the cross-validation.

Representational similarity analysis: model-based regression

We performed multiple linear regression (Eq.1), using the vectorized correlation coefficient RSA matrices described above as the dependent variable (r) and a set of vectorized model RSA matrices as regressors. This analysis estimates the unique contribution of each model RSA matrix to the correlation coefficient RSA matrix. Five model RSA matrices were used for each training period (pre-cue delay matrices depicted in **Fig 5a**; post-cue delay matrices depicted in **Fig S3a**): (1) A matrix representing hemifield spatial attention ('SA') reflecting similarity among stimuli presented on the same visual hemifield, but dissimilarity among stimuli presented on the opposite visual hemifields. Similarity was assigned a regressor value of 1 and dissimilarity was assigned a regressor value of -1. (2) A matrix representing the spatial location of the reference ('Concrete Retro'). Here, distance-dependent similarity was accounted for by computing the distance in visual angle among locations in the same hemifield. The distance in visual angle between two locations was scaled by the maximum distance among any two locations such that a location with itself received a value of 1 and the maximally distant two locations received a value of -1. Between hemifield locations were assigned a value of 0 as these dissimilarities are already captured by the SA regressor. (3) A matrix representing the next location in the sequence ('Concrete Pro'), in which the highest similarity was predicted between the spatial location of the reference and the next spatial location in the sequence. For example, the neural representation of 'L1' would be the most similar to the neural representation of 'L2' reflecting a transformation to the next

location in the sequence. Distances were computed in an identical manner to the 'Concrete Retro' matrix such that the top left and bottom right quadrants of the 'Concrete Pro' matrix were one-element shifts of the corresponding quadrants of the 'Concrete Pro' matrix. This same logic applies to the other matrices described below. (4) A matrix representing the abstract sequence position for the reference stimulus ('Abstract Retro'), wherein the values of the second model RSA matrix ('Concrete Retro') were flipped in the horizontal direction. By doing so, we modeled cross-hemifield generalizability of the abstract sequence position of the past. Note that the ideal matrix corresponding to a retrospective abstract sequence position would be the summation of the 'Concrete Retro' and 'Abstract Retro' matrices as described. However, such a matrix would induce collinearity with the 'Concrete Retro' matrix. Hence, the 'Abstract Retro' matrix as described captures additional variance attributed to cross-hemifield generalization over-and-above within-hemifield distance-dependent similarity. (5) A matrix representing the abstract sequence position for the next location in the sequence ('Abstract Pro'), wherein the values of the third model RSA matrix ('Concrete Pro') were flipped in the horizontal direction. By doing so, we modeled the cross-hemifield generalizability of the abstract sequence position of the upcoming future. As with the 'Abstract Retro' matrix, this matrix captures additional variance attributed to cross-hemifield generalization over-and-above within-hemifield distance-dependent similarity.

$$r \sim \beta_0 + \beta_1 \mathrm{SA} + \beta_2 \mathrm{Concrete\ Retro} + \beta_3 \mathrm{Concrete\ Pro} + \beta_4 \mathrm{Abstract\ Retro} + \beta_5 \mathrm{Abstract\ Pro}$$

Eq.1

For each iteration of the cross-validation procedure, we estimated beta weights for each time point using the equation (Eq. 1). The beta weights obtained from the model-based regression across the iterations were averaged for each participant.

Multidimensional scaling analysis

To visualize the representational geometry of WM, we performed multidimensional scaling (MDS). For each subject, ROI, TR, and location, we obtained the average classifier evidence for each of the 10 spatial locations that were estimated from the decoding analyses where the training and testing the classifier were performed within the same TR.

These classifier evidence were averaged across the time points of the pre-cue and post-cue delays, respectively. This resulted in a 10 x 10 matrix indicating the confusability of each spatial location for each delay period. All participant-level pre- and post-cue matrices were then concatenated (150 X 10 pre- and post-cue matrices), respectively, and used to calculate 10 X 10 correlation distance matrices. These correlation distance matrices were submitted to MDS using the cmdscale function implemented in R. The first three dimensions accounting for 76.6-91.4% variance (**Fig S6a**) were visualized. Each location was projected onto a 3-dimensional space showing the WM representational spaces of the ROIs by time (**Fig 8a**). To statistically quantify the WM representational geometry, we repeated the same procedure separately at the individual participant level. Then, for each participant, we calculated the Euclidean distances between 10 spatial locations based on the 3-dimensional coordinates of spatial locations, resulting in a Euclidean distance matrix (Top row in **Fig 8c**). Since we supposed that the WM representational geometry might reflect either the retinotopic or sequence position space, we estimated a matrix representing each space, using the visual angle distance of spatial locations (Bottom row in **Fig 8c**). In order to measure the similarity between the Euclidean distance and visual angle distance matrices, we calculated the Pearson's correlation between the vectorized upper triangle portion of the Euclidean distance and visual angle distance matrices (excluding the diagonal) and performed Fisher's Z transformation the statistical quantification.

Statistical procedures

Statistical tests on decoding accuracy shown in Figures 2 and 4 were based on permutation testing with 1000 iterations. For each iteration, the classifier was trained and tested with shuffled decoding labels, resulting in 15 null decoding accuracies (1 per participant) for each time point (0, 2, 4, 6, 8, 10, 12, and 14 sec) and each ROI (left/right IPS and VC). These null decoding accuracies were submitted to a one-tailed one-sample t-test against the chance level (one-tailed since decoding accuracy is not expected to be significantly below chance; **Figs 2b and 4b**) and two-way RM ANOVAs (**Figs 2c and 4c**), resulting in null T- or null F-values, which formed the null distributions. P-values were the percentage of values in the null distribution that were greater than or equal to the

observed value obtained from the intact data, resulting in 0.001 as the lower limit for the P-value. We applied this calculation to other permutation statistical tests. (Note that for two-tailed tests, P-values were calculated based on the null distribution of the absolute null T-value and the absolute observed value). For the statistical test on classifier evidence shown in Figure 3, we defined the null distribution at the same time points where the observed classifier evidence was obtained. Then, the observed classifier evidence was compared to the chance level using a one-tailed one-sample t-test, resulting in an intact T-value, which was used to calculate the P-value.

Similar to the above, statistical tests on beta weights obtained from RSA shown in Figures 5, and 6 were based on permutation testing over 1000 iterations. For each iteration, the shuffled correlation coefficient RSA matrix was submitted to model-based regression:

$$shuffled\ r \sim \beta_0 + \beta_1 \mathrm{SA} + \beta_2 \mathrm{Concrete\ Retro} + \beta_3 \mathrm{Concrete\ Pro} + \beta_4 \mathrm{Abstract\ Retro} + \beta_5 \mathrm{Abstract\ Pro} \qquad \textit{Eq.2}$$

For each iteration, we obtained 15 null beta weights (1 per subject) for each time point, each ROI, and each regressor (Concrete Retro, Concrete Pro, Abstract Retro, and Abstract Pro). These null beta weights were submitted to one-tailed one-sample t-tests against a value of 0 (one-tailed because only beta weights above a value of 0 are expected; **Figs 5c and 6a**), two-tailed paired t-tests (**Fig 6a**) and a two-way RM ANOVA (**Fig 6a**), resulting in null T- or null F-values, which formed the null distributions.

The statistical test on the brain-behavior correlation analysis was based on a permutation test with 1000 iterations. For each iteration, we shuffled the participant order of 15 observed averaged beta weights of interest and correlated it with the intact participant order of behavioral data (**Fig 6b**), resulting in the null distribution of the null T-values.

Statistical test on MDS shown in Figure 7d was based on permutation testing with 1000 iterations. Since the output of the cmdscale function in R was used as the coordinates corresponding to WM representations for 10 spatial locations in 3D space, we shuffled

the outputs but kept the label of 10 spatial locations intact for each iteration. Using the shuffled multidimensional 3D coordinates, we calculated the null Euclidean distances between 10 spatial locations, generating a null Euclidean distance matrix. These matrices were then correlated with either the intact retinotopic matrices or sequence position matrices by vectorizing the upper triangular portion of each matrix. For the statistical test, the correlation value was Fisher's Z transformed, resulting in 15 null Fisher's Z values (1 per subject) for each time point, each ROI, and each space (retinotopic and position). These null Fisher's Z values were submitted to two-tailed paired t-tests (**Fig 7d**), and two-way/three-way RM ANOVAs (**Fig 7d**), resulting in the null distribution of the null T- and F-values.

For all statistical tests across multiple time points, multiple comparisons were corrected using false-discovery rate (FDR) [64].

References

1. Baddeley, A., *Working memory.* Science, 1992. **255**(5044): p. 556-9.
2. Curtis, C.E. and M. D'Esposito, *Persistent activity in the prefrontal cortex during working memory.* Trends Cogn Sci, 2003. **7**(9): p. 415-423.
3. D'Esposito, M. and B.R. Postle, *The cognitive neuroscience of working memory.* Annu Rev Psychol, 2015. **66**: p. 115-42.
4. Nee, D.E. and M. D'Esposito, *The Representational Basis of Working Memory.* Curr Top Behav Neurosci, 2018. **37**: p. 213-230.
5. Fuster, J.M., *Prefrontal cortex and the bridging of temporal gaps in the perception-action cycle.* Ann N Y Acad Sci, 1990. **608**: p. 318-29; discussion 330-6.
6. Rainer, G., S.C. Rao, and E.K. Miller, *Prospective coding for objects in primate prefrontal cortex.* Journal of Neuroscience, 1999. **19**(13): p. 5493-5505.
7. Fuster, J.M., *The prefrontal cortex--an update: time is of the essence.* Neuron, 2001. **30**(2): p. 319-33.
8. Nobre, A.C. and M.G. Stokes, *Premembering Experience: A Hierarchy of Time-Scales for Proactive Attention.* Neuron, 2019. **104**(1): p. 132-146.
9. Albers, A.M., et al., *Shared representations for working memory and mental imagery in early visual cortex.* Curr Biol, 2013. **23**(15): p. 1427-31.
10. Bettencourt, K.C. and Y. Xu, *Decoding the content of visual short-term memory under distraction in occipital and parietal areas.* Nat Neurosci, 2016. **19**(1): p. 150-7.
11. Christophel, T.B., et al., *Parietal and early visual cortices encode working memory content across mental transformations.* Neuroimage, 2015. **106**: p. 198-206.
12. Christophel, T.B., M.N. Hebart, and J.D. Haynes, *Decoding the contents of visual short-term memory from human visual and parietal cortex.* J Neurosci, 2012. **32**(38): p. 12983-9.
13. Ester, E.F., T.C. Sprague, and J.T. Serences, *Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory.* Neuron, 2015. **87**(4): p. 893-905.
14. Harrison, S.A. and F. Tong, *Decoding reveals the contents of visual working memory in early visual areas.* Nature, 2009. **458**(7238): p. 632-5.
15. Li, H.H. and C.E. Curtis, *Neural population dynamics of human working memory.* Curr Biol, 2023. **33**(17): p. 3775-3784 e4.
16. Lorenc, E.S., et al., *Flexible Coding of Visual Working Memory Representations during Distraction.* J Neurosci, 2018. **38**(23): p. 5267-5276.
17. Rademaker, R.L., C. Chunharas, and J.T. Serences, *Coexisting representations of sensory and mnemonic information in human visual cortex.* Nat Neurosci, 2019. **22**(8): p. 1336-1344.
18. Serences, J.T., et al., *Stimulus-specific delay activity in human primary visual cortex.* Psychol Sci, 2009. **20**(2): p. 207-14.
19. Sprague, T.C., E.F. Ester, and J.T. Serences, *Reconstructions of information in visual spatial working memory degrade with memory load.* Curr Biol, 2014. **24**(18): p. 2174-2180.

20. Sprague, T.C., E.F. Ester, and J.T. Serences, *Restoring Latent Visual Working Memory Representations in Human Cortex.* Neuron, 2016. **91**(3): p. 694-707.
21. Sprague, T.C. and J.T. Serences, *Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices.* Nat Neurosci, 2013. **16**(12): p. 1879-87.
22. Pasternak, T. and M.W. Greenlee, *Working memory in primate sensory systems.* Nat Rev Neurosci, 2005. **6**(2): p. 97-107.
23. Scimeca, J.M., A. Kiyonaga, and M. D'Esposito, *Reaffirming the Sensory Recruitment Account of Working Memory.* Trends in Cognitive Sciences, 2018. **22**(3): p. 190-192.
24. Serences, J.T., *Neural mechanisms of information storage in visual short-term memory.* Vision Res, 2016. **128**: p. 53-67.
25. Sreenivasan, K.K. and M. D'Esposito, *The what, where and how of delay activity.* Nature Reviews Neuroscience, 2019. **20**(8): p. 466-481.
26. Badre, D., A.S. Kayser, and M. D'Esposito, *Frontal cortex and the discovery of abstract action rules.* Neuron, 2010. **66**(2): p. 315-26.
27. Duan, Z.Y. and C.E. Curtis, *Visual working memories are abstractions of percepts.* Elife, 2024. **13**.
28. Kwak, Y. and C.E. Curtis, *Unveiling the abstract format of mnemonic representations.* Neuron, 2022. **110**(11): p. 1822-1828 e5.
29. Boettcher, S.E.P., et al., *Output planning at the input stage in visual working memory.* Science Advances, 2021. **7**(13).
30. Nasrawi, R. and F. van Ede, *Planning the Potential Future during Multi-item Visual Working Memory.* Journal of Cognitive Neuroscience, 2022. **34**(8): p. 1534-1546.
31. Shushruth, S., A. Zylberberg, and M.N. Shadlen, *Sequential sampling from memory underlies action selection during abstract decision-making.* Current Biology, 2022. **32**(9): p. 1949-+.
32. van Ede, F., et al., *Concurrent visual and motor selection during visual working memory guided action.* Nature Neuroscience, 2019. **22**(3): p. 477-+.
33. Mesulam, M.M., *From sensation to cognition.* Brain, 1998. **121 (Pt 6)**: p. 1013-52.
34. Murray, J.D., et al., *Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex.* Proc Natl Acad Sci U S A, 2017. **114**(2): p. 394-399.
35. Badre, D., et al., *The dimensionality of neural representations for control.* Curr Opin Behav Sci, 2021. **38**: p. 20-28.
36. Badre, D. and M. D'Esposito, *Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex.* Journal of Cognitive Neuroscience, 2007. **19**(12): p. 2082-2099.
37. Bernardi, S., et al., *The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex.* Cell, 2020. **183**(4): p. 954-967 e21.
38. Nee, D.E. and M. D'Esposito, *The hierarchical organization of the lateral prefrontal cortex.* Elife, 2016. **5**.

39. King, J.R. and S. Dehaene, *Characterizing the dynamics of mental representations: the temporal generalization method.* Trends Cogn Sci, 2014. **18**(4): p. 203-10.
40. Panichello, M.F. and T.J. Buschman, *Shared mechanisms underlie the control of working memory and attention.* Nature, 2021. **592**(7855): p. 601-605.
41. Piwek, E.P., M.G. Stokes, and C. Summerfield, *A recurrent neural network model of prefrontal brain activity during a working memory task.* Plos Computational Biology, 2023. **19**(10).
42. Wan, Q., et al., *Representing Context and Priority in Working Memory.* Journal of Cognitive Neuroscience, 2024. **36**(7): p. 1374-1394.
43. Yu, Q., C. Teng, and B.R. Postle, *Different states of priority recruit different neural representations in visual working memory.* PLoS Biol, 2020. **18**(6): p. e3000769.
44. Lorenc, E.S., et al., *Dissociable neural mechanisms underlie currently-relevant, future-relevant, and discarded working memory representations.* Sci Rep, 2020. **10**(1): p. 11195.
45. LaRocque, J.J., et al., *Within-Category Decoding of Information in Different Attentional States in Short-Term Memory.* Cereb Cortex, 2017. **27**(10): p. 4881-4890.
46. Lewis-Peacock, J.A., et al., *Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention.* Journal of Cognitive Neuroscience, 2012. **24**(1): p. 61-79.
47. van Ede, F., J. Deden, and A.C. Nobre, *Looking ahead in working memory to guide sequential behaviour.* Current Biology, 2021. **31**(12): p. R779-R780.
48. Gunseli, E., et al., *Encoded and updated spatial working memories share a common representational format in alpha activity.* iScience, 2024. **27**(2): p. 108963.
49. Liu, B.W., Z.S. Alexopoulou, and F. van Ede, *Jointly looking to the past and the future in visual working memory.* Elife, 2024. **12**.
50. Lewis-Peacock, J.A. and B.R. Postle, *Temporary activation of long-term memory supports working memory.* J Neurosci, 2008. **28**(35): p. 8765-71.
51. Xu, Y.D., *Parietal-driven visual working memory representation in occipito-temporal cortex.* Current Biology, 2023. **33**(20): p. 4516-+.
52. Myers, N.E., M.G. Stokes, and A.C. Nobre, *Prioritizing Information during Working Memory: Beyond Sustained Internal Attention.* Trends Cogn Sci, 2017. **21**(6): p. 449-461.
53. van Ede, F. and A.C. Nobre, *Turning Attention Inside Out: How Working Memory Serves Behavior.* Annual Review of Psychology, 2023. **74**: p. 137-165.
54. Wang, L.P., et al., *Representation of spatial sequences using nested rules in human prefrontal cortex.* Neuroimage, 2019. **186**: p. 245-255.
55. Xie, Y., et al., *Geometry of sequence working memory in macaque prefrontal cortex.* Science, 2022. **375**(6581): p. 632-+.
56. Mishkin, M., L.G. Ungerleider, and K.A. Macko, *Object vision and spatial vision: two cortical pathways.* Trends in neurosciences, 1983. **6**: p. 414-417.
57. Andersson, J.L., et al., *Modeling geometric deformations in EPI time series.* Neuroimage, 2001. **13**(5): p. 903-19.

58. Lund, T.E., et al., *Motion or activity: their role in intra-and inter-subject variation in fMRI.* Neuroimage, 2005. **26**(3): p. 960-964.
59. Ashburner, J. and K. Friston, *Multimodal image coregistration and partitioning--a unified framework.* Neuroimage, 1997. **6**(3): p. 209-17.
60. Wang, L., et al., *Probabilistic Maps of Visual Topography in Human Cortex.* Cerebral Cortex, 2015. **25**(10): p. 3911-3931.
61. Kumar, M., et al., *BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis.* PLoS Comput Biol, 2020. **16**(1): p. e1007549.
62. Pedregosa, F., *Scikit‐learn: Machine learning in python Fabian.* Journal of machine learning research, 2011. **12**: p. 2825.
63. Kriegeskorte, N., M. Mur, and P. Bandettini, *Representational similarity analysis - connecting the branches of systems neuroscience.* Front Syst Neurosci, 2008. **2**: p. 4.
64. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society Series B-Statistical Methodology, 1995. **57**(1): p. 289-300.

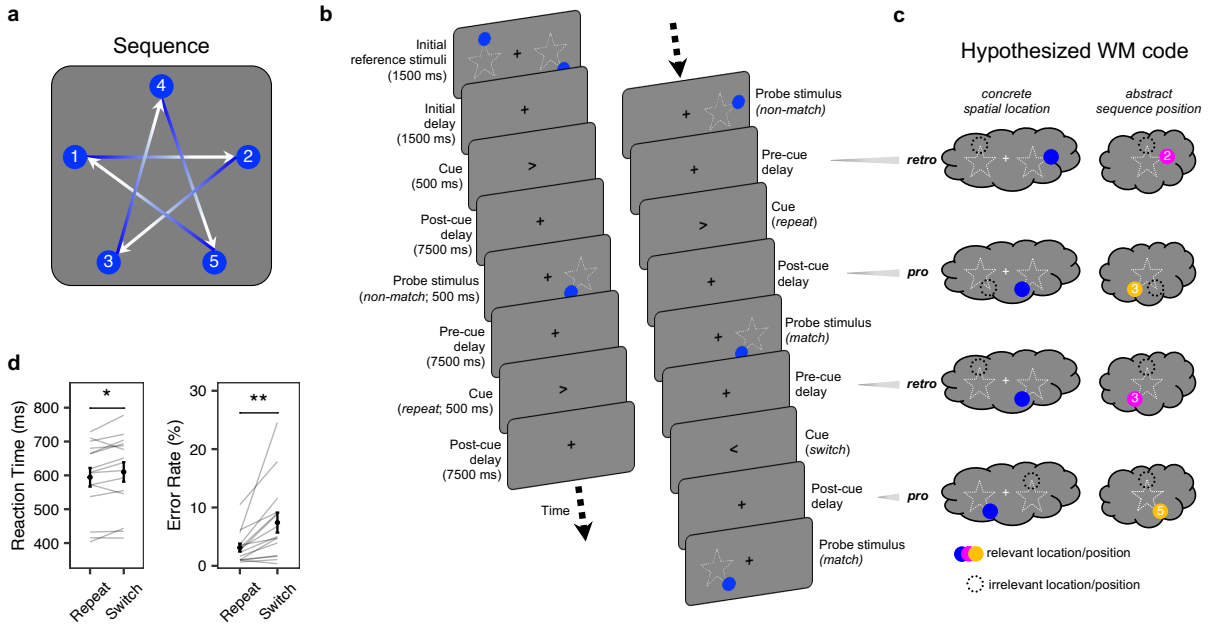

Fig 1. Procedures and behavioral performance.
(a) The sequence consisted of five points forming a star. Arrows indicate the sequence ordering such that position 2 followed position 1, position 3 followed position 2, etc. Position 5 transitioned back to position 1 such that the sequence was circular.
(b) Task procedure. Each trial consisted of a pre-cue delay, a cue, and a post-cue delay. At the start of each run, initial spatial locations were presented in the left and right visual hemifields. After an initial delay, a cue indicated the task-relevant visual hemifield wherein a forthcoming probe would be presented following a post-cue delay. Participants determined whether the spatial location of the probe followed the spatial location of the last item that appeared in the same hemifield. Half of the probes were sequence matches (e.g., position 2 following position 1) while half were non-matches (e.g., position 3 following position 1). The star trace is used for illustrative purpose and did not appear in the actual task.
(c) We hypothesized that during the pre-cue delay, the spatial location of the just-probed location (concrete, retrospective code; hemifield-specific) and its sequence position (abstract, retrospective code; hemifield-general) would be maintained in working memory. During the post-cue delay, we hypothesized that retrospective codes would be transformed towards future expectations. Specifically, the next sequence position (abstract, prospective code; hemifield-general) would lead to a prediction of the next spatial location in the sequence (concrete, prospective code; hemifield-specific). For completeness, we depict analogous coding possibilities for the irrelevant (uncued) locations/positions although these were not of central interest.
(d) Mean reaction time and error rates. Behavioral performance was better when the cue indicated the same visual hemifield as the previous trial ('Repeat') relative to when the cue indicated the other visual hemifield ('Switch'). Each line on the bar graph represents each participant. Bold dots indicate the mean across participants with error bars representing ± 1 standard error of mean. *$p < 0.05$, **$p < 0.01$.

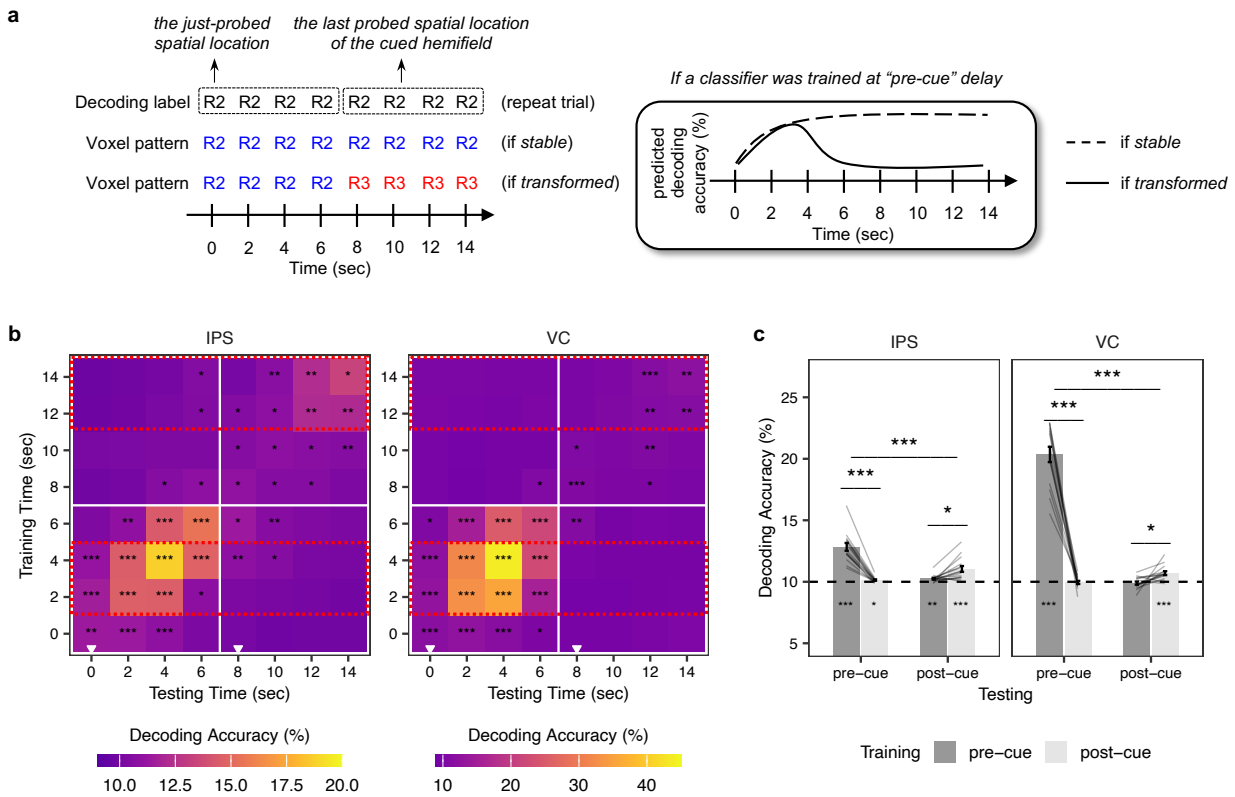

Fig 2. Working memory content maintained during the post-cue delay differs from that maintained during the pre-cue delay.
(a) Each time point was labeled based on the just-probed location (pre-cue) and/or last probed location in the cued hemifield (post-cue). Whether the WM remained stable or was transformed predicts different patterns of decoding accuracy across pre- and post-cue delay periods.
(b) Temporal generalization decoding. Classifiers were trained on a given time point and tested across time points. Decoding performance was compared against the chance level of 10% and corrected for multiple comparisons using false discovery rate (FDR). The red dotted boxes on the graph indicate the time points of maximal decoding for a given period which were used for training in subsequent analyses (**Fig 3**). *q < 0.05 and **q < 0.01 (FDR corrected).
(c) Average of decoding accuracies in each delay period. The horizontal dotted line indicates the chance level of 10%. Each line on the bar graph represents each participant. *p < 0.05, **p < 0.01, ***p = 0.001. Error bars represent ±1 standard error of mean.

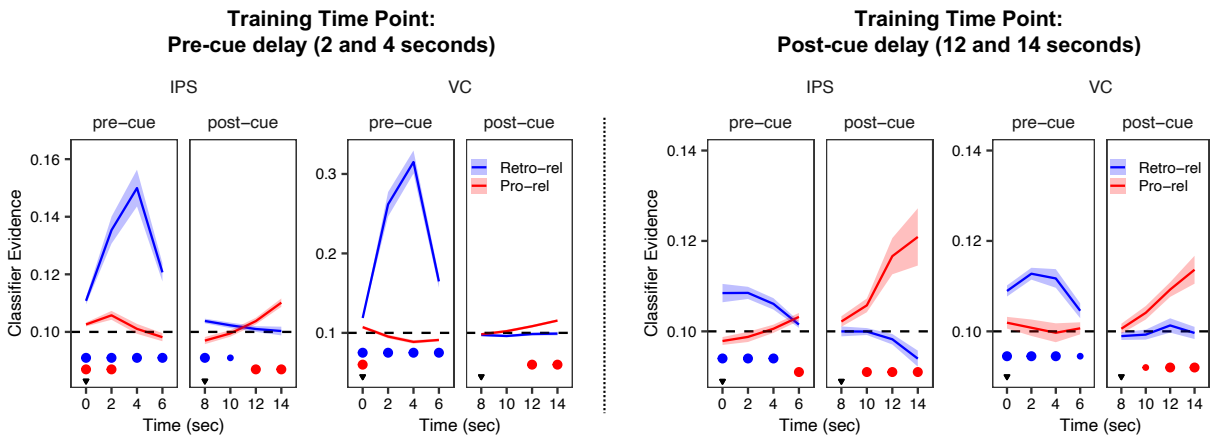

Fig 3. Tracking retrospective and prospective codes over time. Separate classifiers were trained using data from the pre-cue (2 and 4 seconds) and post-cue (12 and 14 seconds) delays (see red dotted boxes in **Fig 2b**). Retrospective-relevant ('Retro-rel') refers to the classifier evidence corresponding to the spatial location of the just-probed item. Prospective-relevant ('Pro-rel') refers to the classifier evidence corresponding to the spatial location following the reference (see **Fig 1c**). The significance at each time point is denoted by small and medium dots, representing $q < 0.05$ and $q < 0.01$, respectively (FDR corrected). The inverted triangles denote the time points of the presentation of the probe and the cue, respectively.

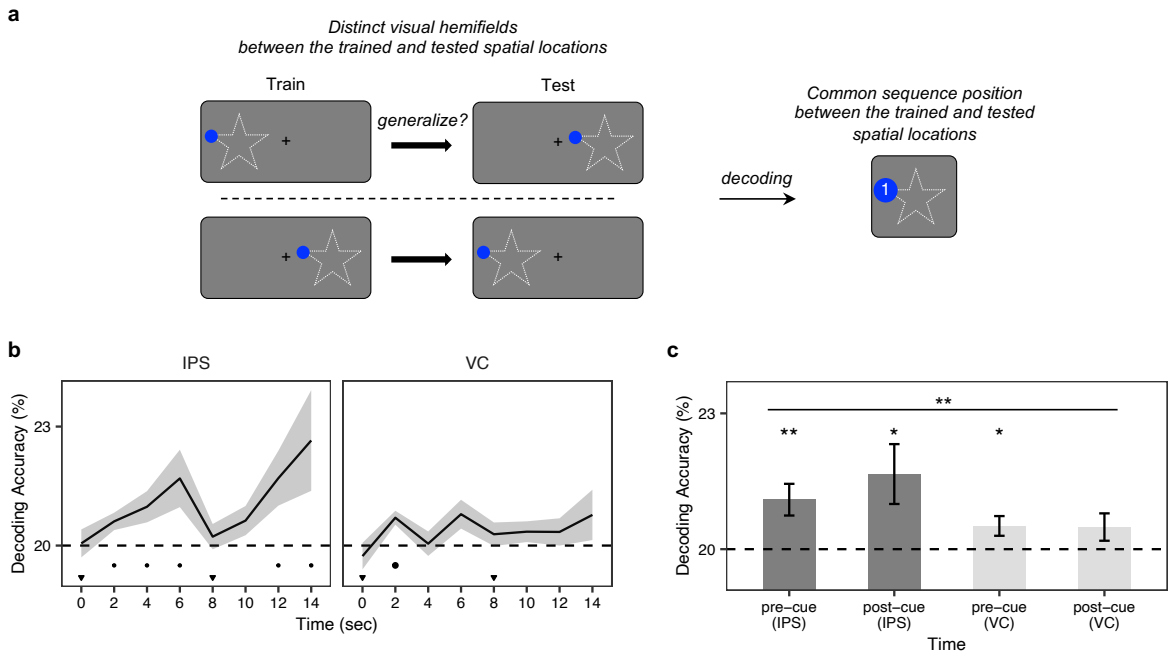

Fig 4. Decoding abstract sequence positions.
(a) Abstract coding was operationally defined based on generalization of decoding across visual hemifields.
(b) The result of the cross-visual hemifield decoding analysis, in which training and testing the classifier were performed on different visual hemifields, respectively. Decoding performance was compared against the chance level of 20% and corrected for multiple comparisons using false discovery rate (FDR). The significance at each time point is denoted by small, and medium dots, representing $q < 0.05$, and $q < 0.01$, respectively. The shaded regions represent ± 1 standard error of mean. The inverted triangles in black denote the time points of the presentation of the probe stimulus and the cue, respectively.
(c) Average of decoding accuracies from 'pre-cue' (2 sec to 6 sec) and 'post-cue' (10 sec to 14 sec) period in each ROI. The horizontal dashed line indicates the chance level of 20%. Each dot on the bar graph represents a participant. $^*p < 0.05$, $^{**}p < 0.01$. Error bars represent ± 1 standard error of mean.

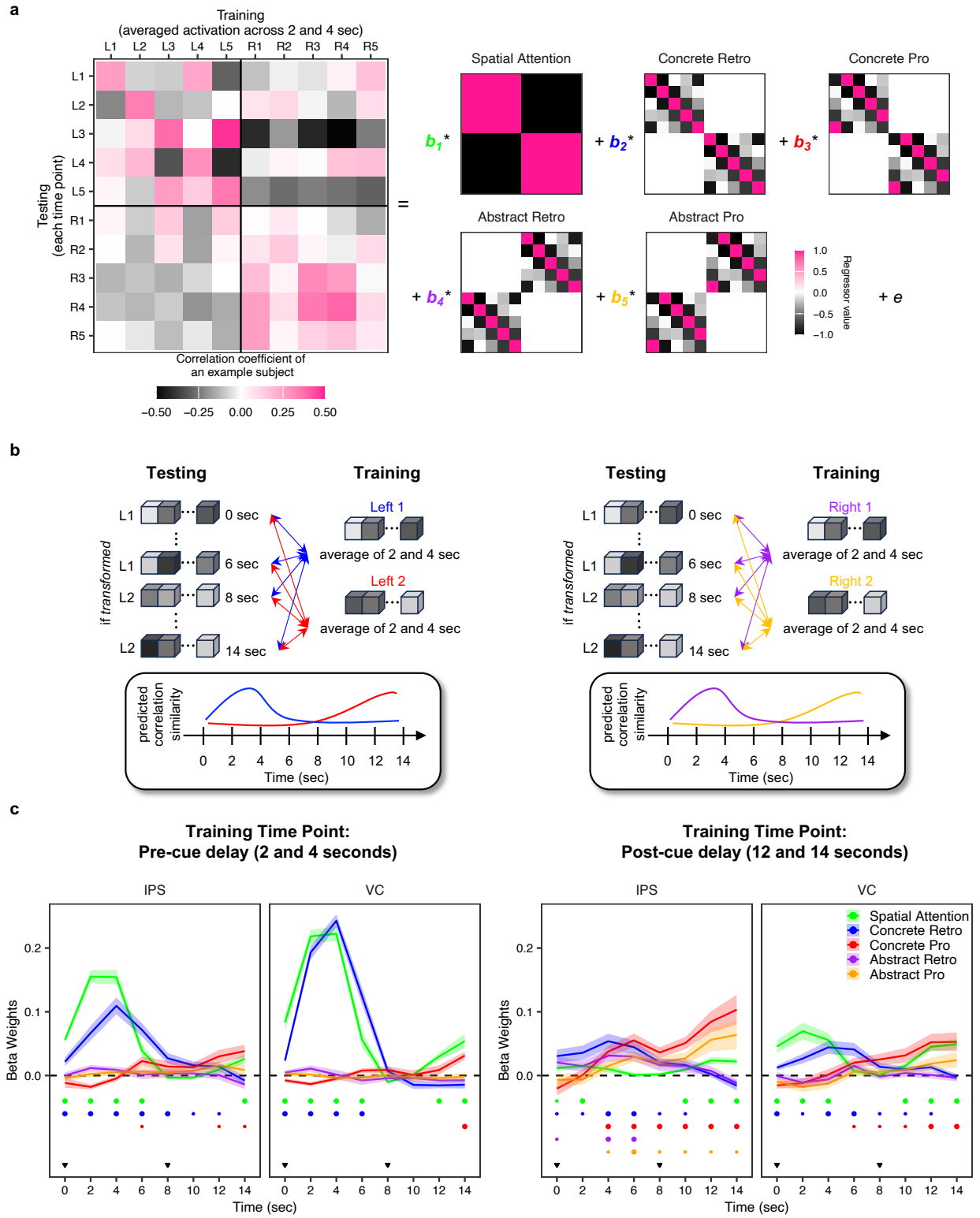

Fig 5. Concrete and abstract codes of working memory are transformed over time.

(a) The left depicts an example representational similarity matrix formed by correlating training patterns of each location with testing patterns of each location. These matrices

were submitted to multiple linear regression using regressors denoting idealized matrices of spatial attention, concrete retrospective coding, concrete prospective coding, abstract retrospective coding, and abstract prospective coding. We estimated the beta weight for each regressor at each time point.
(b) Schematic of the predicted correlation similarities representing a transition between concrete retrospective to concrete prospective coding (left) and abstract retrospective to abstract prospective coding (right). (See also **Fig S5** for the analogous regressors for the post-cue delay trained data).
(c) The estimated beta weights for each regressor at each time point. The beta weights were compared against a value of 0 and corrected for multiple comparisons using false discovery rate. The significance at each time point is denoted by small and medium dots, representing $q < 0.05$ and $q < 0.01$, respectively (FDR corrected). The shaded regions represent ± 1 standard error of mean. The inverted triangles in black denote the time points of the presentation of the probe stimulus and the cue, respectively.

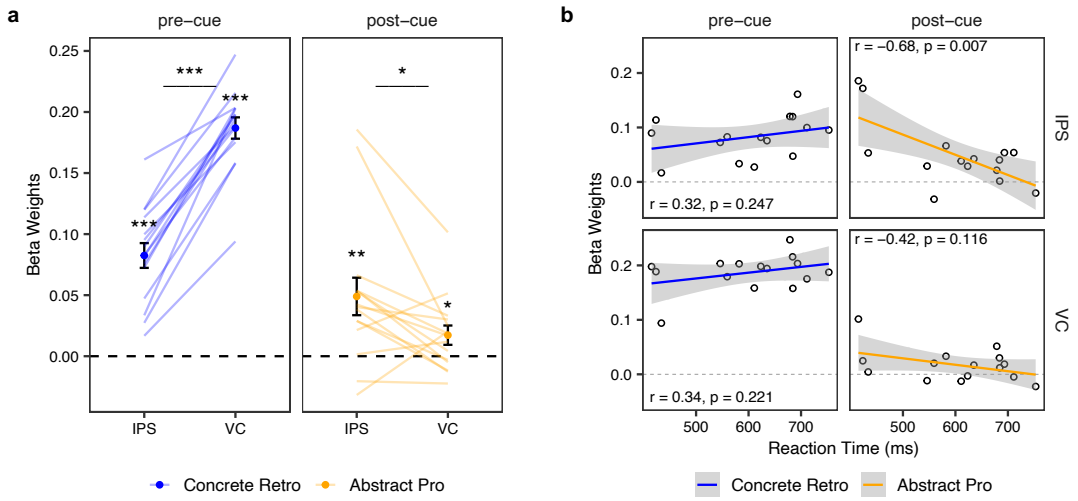

Fig 6. Different roles of IPS and VC in representing WM codes.
(a) Contrasting concrete retrospective (Concrete Retro) and abstract prospective (Abstract Pro) codes in the intra-parietal sulcus (IPS) and visual cortex (VC). The concrete retrospective code was formed by averaging beta weights during the 'pre-cue' period (2 sec to 6 sec) for each ROI using the pre-cue trained data (blue). The abstract prospective code was formed by averaging beta weights during the 'post-cue' period (10 sec to 14 sec) for each ROI using the post-cue training period (orange). Each line represents each participant. Bold color dots indicate the mean across participants with error bars representing ± 1 standard error of mean. *$p < 0.05$, **$p < 0.01$, ***$p = 0.001$.
(b) Correlations between the strengths of concrete retrospective (blue) and abstract prospective (orange) and RT for each ROI. Each dot represents each participant. The shaded regions represent 95% confidence interval of the correlation.

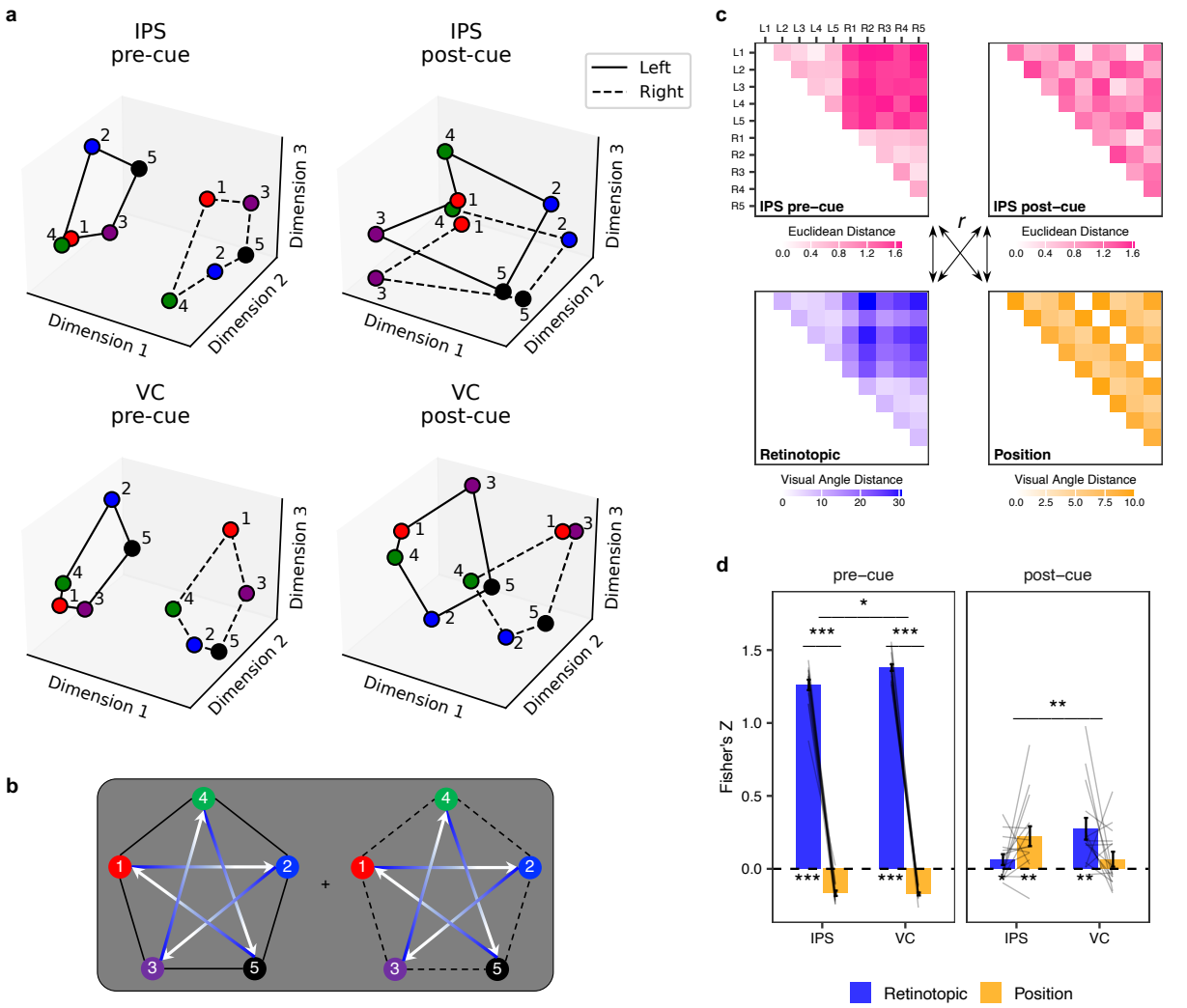

Fig 7. Representational geometry of WM.
(a) Each of the 10 spatial locations is plotted as a function of the first three dimensions determined by multidimensional scaling (MDS). Spatial locations are separately color coded by sequence position. Spatial locations in the left and right visual hemifields are connected by solid and dotted black lines, respectively. (b) Schematic illustration of the spatial locations. (c) Group-level Euclidean distance in the multidimensional space between spatial locations in the IPS pre- and post-cue (top row). Visual angle distance between spatial locations (bottom row). The visual angle distance was estimated considering retinotopic (bottom left) and sequence position geometries (bottom right). (d) The result of correlation between Euclidean distance and visual angle distance. Each line on the bar graph represents each participant. *p < 0.05, **p < 0.01, ***p = 0.001. Error bars represent ±1 standard error of mean.