Title

Long-term learning transforms prefrontal cortex representations during working memory.

Permalink

https://escholarship.org/uc/item/86c5r2jb

Journal

Neuron, 110(22)

Authors

Miller, Jacob

Tambini, Arielle

Kiyonaga, Anastasia

et al.

Publication Date

2022-11-16

DOI

10.1016/j.neuron.2022.09.019

Peer reviewed

Published in final edited form as:
Neuron. 2022 November 16; 110(22): 3805–3819.e6. doi:10.1016/j.neuron.2022.09.019.

Long-term learning transforms prefrontal cortex representations during working memory

Jacob A. Miller[1,6,*], **Arielle Tambini**[2], **Anastasia Kiyonaga**[3], **Mark D'Esposito**[4,5]

[1]Wu Tsai Institute, Department of Psychiatry, Yale University, New Haven, CT, USA

[2]Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA

[3]Department of Cognitive Science, University of California, San Diego, CA, USA

[4]Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

[5]Department of Psychology, University of California, Berkeley, CA, USA

[6]Lead contact

SUMMARY

The role of the lateral prefrontal cortex (lPFC) in working memory (WM) is debated. Non-human primate (NHP) electrophysiology shows that the lPFC stores WM representations, but human neuroimaging suggests that the lPFC controls WM content in sensory cortices. These accounts are confounded by differences in task training and stimulus exposure. We tested whether long-term training alters lPFC function by densely sampling WM activity using functional MRI. Over 3 months, participants trained on both a WM and serial reaction time (SRT) task, wherein fractal stimuli were embedded within sequences. WM performance improved for trained (but not novel) fractals and, neurally, delay activity increased in distributed lPFC voxels across learning. Item-level WM representations became detectable within lPFC patterns, and lPFC activity reflected sequence relationships from the SRT task. These findings demonstrate that human lPFC develops stimulus-selective responses with learning, and WM representations are shaped by long-term experience, which could reconcile competing accounts of WM functioning.

In brief

Miller et al. densely sampled brain activity with human neuroimaging during working memory across months of learning. Long-term training altered the role of the prefrontal cortex, which developed representations for specific stimuli and associations learned over time. Working memory is shaped by long-term experience, which may help resolve competing accounts of prefrontal functioning.

[*]Correspondence: j.a.miller@yale.edu.

AUTHOR CONTRIBUTIONS
J.A.M., A.K., and A.T. developed the study concept and design with input from M.D. J.A.M., A.K., and A.T. collected and analyzed the data. J.A.M., A.K., and A.T. drafted the manuscript, and M.D. provided feedback and revisions. All authors approved the final version of the manuscript for submission.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2022.09.019.

INTRODUCTION

The lateral prefrontal cortex (lPFC) is considered critical for working memory (WM) across human and animal models (Funahashi et al., 1989; Goldman-Rakic, 1995; Leavitt et al., 2017; Miller et al., 2018; Sreenivasan et al., 2014). However, there is ongoing debate regarding the specific role that lPFC activity plays in successful WM (Christophel et al., 2017; Curtis and Sprague, 2021; Lara and Wallis, 2015; Mackey et al., 2016). Non-human primate (NHP) electrophysiology research typically finds that the lPFC maintains feature-specific WM content (Constantinidis et al., 2018; Funahashi et al., 1989; Fuster and Alexander, 1971; Goldman-Rakic, 1995; Miller et al., 2018; Romo et al., 1999). Human neuroimaging suggests that lPFC activity serves control functions over WM, with feature-specific content stored in sensory cortices (D'Esposito and Postle, 2015; Harrison and Tong, 2009; Riggall and Postle, 2012; Serences, 2016). However, these seemingly incompatible accounts are confounded by differences in species, measurement granularity, and the amount of task training.

One possibility is that different indices of neural activity, across measurement scales, may support distinct conclusions about the cortical substrates for WM. That is, NHP studies typically record finer resolution single-unit neuronal activity compared with the millimeter scale of blood-oxygen-level-dependent functional MRI (BOLD fMRI) (Mukamel et al., 2005; Park et al., 2017). Discrepancies may emerge if stimulus-specific WM content is represented in the human lPFC but is undetectable at the coarser resolution of BOLD fMRI—for instance, via spiking patterns across spatially intermixed neural populations. The organization and spread of activity in sensory areas better matches the spatial resolution of BOLD fMRI, which may also reflect local field potentials from top-down modulation in the absence of local spiking (Leavitt et al., 2017; Lorenc and Sreenivasan, 2021; Mendoza-Halliday et al., 2014; Serences, 2016). However, in some cases, stimulus-specific WM delay activity has been detected in human frontal cortex (Ester et al., 2015; Lee et al., 2013) or NHP sensory regions (Mendoza-Halliday et al., 2014; Supèr et al., 2001), high-lighting the need to identify which factors truly drive observed differences in findings across studies.

In addition to differences in recording techniques between human and NHP studies, NHPs typically perform orders of magnitude more task trials, over months of training, before neural recordings occur (Berger et al., 2018; Birman and Gardner, 2016; Sarma et al., 2016). Humans typically complete only a few minutes of task practice prior to fMRI scanning. Differences observed in neural WM substrates across species may therefore be driven by long-term learning influences from extensive task and stimulus experience. In fact, the few studies that recorded from NHPs before and after WM training found plasticity in the form of increases in the magnitude of WM delay activity and the strength of item-level stimulus representations in the anterior lPFC (Dang et al., 2021; Meyer et al., 2011; Tang et al., 2019; Riley et al., 2018; Sarma et al., 2016). The human lPFC may like-wise represent item-level information in WM, depending on the level of prior training. However, the typical timeline of fMRI research has limited our ability to directly test the hypothesis that WM representations change with long-term learning.

The brain regions and neural mechanisms for WM are classically considered separate from long-term memory (LTM) systems (Squire and Zola-Morgan, 1991; Warrington and Shallice, 1969; Wickelgren, 1996). However, some WM theories predict that learned associations or semantic links between items should be reflected during WM maintenance (LaRocque et al., 2014; Oberauer, 2009), and growing evidence suggests common neural machinery between WM and LTM (Beukers et al., 2021; Fukuda and Woodman, 2017; Hoskin et al., 2019; Lewis-Peacock and Norman, 2014; Nee and Jonides, 2011; Ranganath and Blumenfeld, 2005; Ranganath et al., 2003; Yonelinas, 2013). In some cases, WM capacity is greater for stimuli with extensive exposure (Aspet al., 2021; Brady et al., 2016; Jackson and Raymond, 2008; Xie and Zhang, 2017), suggesting that WM and supporting neural mechanisms may change with stimulus experience.

Here, we examined the possibility that long-term learning transforms human lPFC WM activity. We asked whether stimulus selectivity emerges in the human lPFC as a function of training, akin to the stimulus-specific WM activity patterns typically found in NHP studies. To do so, three human participants each completed over 20 sessions of whole-brain fMRI along with at-home training across 3 months. During this time, participants repeatedly performed a delayed recognition WM task and a sequence learning task, both of which employed a set of 18 novel fractal stimuli that were unique to each participant. First, we asked whether the lPFC activity during the WM delay period changed in magnitude across learning. Widespread decreases in lPFC activity could suggest more automatic task processing with training. Activity increases, however, could suggest greater selectivity for the repeated task structure or individual WM stimuli (Constantinidis et al., 2018; Curtis and Sprague, 2021; Murray et al., 2017). We then tested whether representations of individual stimuli or associative structures emerged in multivariate WM activity patterns across learning. If item-level lPFC activity patterns develop over time, it would suggest that differences in participant training may explain discrepant accounts of the lPFC as a source of control over WM (from human studies) versus WM content storage (from single-unit NHP studies). Alternatively, long-term learning may enhance sensory representations of WM content but induce no changes in the lPFC, suggesting that differences in lPFC versus sensory-based WM storage models are driven by other factors than long-term learning. Finally, to understand how WM representations are shaped by associative learning, we asked whether associations between stimuli learned outside of the WM task were reflected in WM activity patterns. To preview the results, long-term learning changed the distribution and stimulus information content of the lPFC WM delay activity, indicating that WM maintenance mechanisms may be flexible to the extent and nature of prior experience with the WM information. These results suggest that differences in the extent of training across species may masquerade as differences in lPFC function.

RESULTS

Training improves WM performance for trained, but not novel, stimuli

To determine how long-term learning influences cortical activity patterns underlying WM maintenance, we trained three human participants on a unique set of fractal stimuli (Figure 1A) over three months. These stimuli had no preexisting meaning and have been used to characterize the influence of long-term associative learning on neural selectivity (Ghazizadeh et al., 2018; Kim et al., 2015; Sakai and Miyashita, 1991). These complex stimuli were chosen to extend the time course of learning and to necessitate a detailed item representation for successful performance. During the study period, each participant completed 17 scanning (fMRI) sessions along with at-home behavioral training sessions multiple times per week (Figure 1B; STAR Methods). During each fMRI session, participants performed two primary tasks, a serial reaction time (SRT) task followed by a WM task (Figures 1C and 1D). The WM task entailed a single-item delayed recognition test wherein the WM sample was either a fractal stimulus from the training set or a novel fractal that appeared only during that session. The first time each participant saw their unique set of 18 stimuli was during the first scanning session. The SRT task used the same 18 trained fractal stimuli, and 12 of the stimuli were embedded in high-probability sequences (Figure 1C). The sequences were unrelated to the goals of the WM task (which was always to remember a single item), but we took advantage of the sequence structure to analyze whether item-level WM representations reflected associations from the SRT task.

Across learning, behavior in the WM task improved for trained (but not novel) stimuli (Figure 1E). Mean WM probe accuracy (% correct) for trained stimuli improved by 23% across the 17 sessions. To characterize the change in WM performance over time, we used a fixed-effects logistic model that can flexibly detect changes in learning over time, estimate *when* these changes are most prominent (inflection point), and adapt to different rates of learning (Figure 2A). Any significant change over time was assessed by correlating the predicted logistic model values with the actual data using cross-validation (STAR Methods; Figure 2A). There was a significant increase in WM accuracy for trained stimuli ($r = 0.77$, $p < 0.001$), and no reliable change for novel stimuli ($r = 0.17$, $p = 0.07$). This increase in WM accuracy for trained versus novel stimuli was confirmed by testing for an interaction between session number (1 → 17; mean-centered) and stimulus category (*trained* versus *novel*) with a fixed-effects linear model ($t(96) = 2.76$, $p = 0.007$).

A complementary pattern emerged for WM probe response time (RT).There was a significant interaction between session number and stimulus category ($t(96) = -4.4$, $p < 0.001$), which was driven by faster responses for trained stimuli over time ($r = -0.54$, $p < 0.001$), with no significant change for novel stimuli ($r = 0.22$, $p = 0.06$). The subsequent analyses use fixed-effects logistic models to flexibly detect changes that occur at different times and rates (see Figure 2A for schematic), but all results generalize to a linear framework.

In parallel with the WM task, participants also learned associations between stimuli that were part of regularly occurring sequences in the SRT task. Reliable associative learning across training was shown by reduced response times for intact sequences in the SRT task for all participants (Figure S1).

Divergent changes in mean WM delay activity within the dorsal PFC

To determine whether lPFC activity changes across learning, we split the lPFC into six bilateral regions of interest (ROIs) along rostral-caudal and dorsal-ventral axes (Figure 2B). We tested for evidence of broad changes in mean WM delay activity over time by considering two groups of voxels within each ROI. First, we examined whether peak activation in the WM delay period changed across sessions (Curtis and Sprague, 2021). To do this, we thresholded WM delay activity maps (collapsed across all delay lengths) for each participant and session at t > 2.5 and determined whether peak activation levels (*beta*-coefficient) changed over training (Figure 2C, left). Second, we analyzed the mean activity of all voxels across each ROI, without any thresholding, to ask whether there are changes across an entire cortical region. Changes in highly active voxels are sensitive to the magnitude of peak activity, but the precise location of highly activated voxels can shift from session to session.

The magnitude of WM delay activity changed across training in one lPFC area. The peak WM delay period activity in dorsal rostral PFC decreased across sessions (r = −0.35, p = 0.007 [false-discovery rate (FDR)-corrected, p = 0.039]; Figure 2C, left), whereas the mean activity for all voxels in this area did not significantly change (r = 0.28, p = 0.02 [FDR-corrected, p = 0.13], this model failed to converge with cross-validation, Figure 2C, right). No other ROIs showed training-related changes in either peak WM delay activity or mean across all voxels (FDR-corrected p values > 0.1; Figure S2). However, this approach may obscure divergent changes that occur within specific populations of voxels with learning. We next used a voxel-wise regression to directly test whether individual voxels increased or decreased activity over time.

More PFC territory is recruited for WM delay activity across learning

Populations of voxels involved in WM maintenance may change their activity over training, as the stimuli and task become increasingly well-learned. For example, WM processing could become more “efficient” by recruiting less cortical territory. Alternatively, more cortical territory could be engaged in representing and processing newly learned stimuli and task dimensions. To test these different predictions, for each voxel, we assessed the relationship between WM delay activity and training session with a logistic model. We tested whether a meaningful proportion of voxels within each frontal ROI show systematic changes in activity over training compared with chance (STAR Methods). A schematic of this approach is shown in Figure 3A, allowing us to test whether populations of voxels show divergent increases or decreases in WM delay activity with training—information that is lost when averaging across voxels.

In three lPFC ROIs, a distributed group of voxels increased in WM delay activity across the 17 sessions compared with chance (Figure 3B; dorsal rostral: p = 0.003 [FDR-corrected, p = 0.018], dorsal mid-lateral: p = 0.003 [FDR-corrected, p = 0.018], ventral caudal: p = 0.01 [FDR-corrected, p = 0.044]; permutation tests). Ventral rostral PFC also had a group voxels with increased activity (p = 0.04), but this did not survive correction (FDR-corrected, p = 0.11). The dorsal mid-lateral and ventral caudal PFC showed the largest percentage of voxels with increasing WM delay activity over months of training (~25% of voxels). Only in the dorsal caudal PFC ROI did a distinct group of voxels show decreased activity (p = 0.03), but this did not survive correction (FDR-corrected, p = 0.09). The topography of activation changes over time and mean activity are shown in Figures 3 and Figure S5, and are available on NeuroVault (neurovault.org/collections/12687/). These observed changes across the lPFC were specific to the WM delay period, as the encoding (sample) period instead showed widespread decreases in activity with training in the lPFC (Figure S3).

In summary, repeated task and stimulus exposure was most commonly associated with increased delay activity in a distributed group of lPFC voxels, suggesting that these areas become more involved in WM maintenance with training. However, this increased activity may stem from developing selectivity for individual stimuli over time, or a nonspecific WM maintenance process that conveys no item-level information content.

Representational similarity emerges for individual items, stimulus category, and sequence category

We next tested whether the multivariate activity patterns across voxels develop stimulus specificity over time. We used a pattern similarity analysis framework to test whether specific representations appear in multi-voxel patterns of delay activity across training. These analyses were designed to test directed hypotheses about training-related changes in item- and category-level representations. We therefore contrasted specific stimulus-pairs with each other to capture various levels of representation: individual items, training category (trained versus novel items), and sequence membership category. We estimated the similarity of representations across individual stimuli (matrices shown in Figure 4; STAR Methods) by computing correlations between WM delay period activity patterns for each stimulus. We then created several models to capture hypothesized levels of representational information (item-level, category-level, sequence category) and tested how well the observed similarity patterns matched the idealized models, producing a measure of representational “pattern strength” for each ROI in each session (STAR Methods; Figure 4A). We then tested whether pattern strength for each model changed across sessions, using the same logistic modeling as in prior analyses. To determine whether any pattern similarity effects were specific to the lPFC or also reflected in sensory areas, we examined patterns from early visual cortex (V1–V4) and the lateral occipital complex (LOC) (STAR Methods).

First, we tested whether distinct representations of individual WM items emerged across training in lPFC or visual ROIs. We operationalized an item-level model for individual stimulus representations by testing for greater within-item pattern similarity (maintenance of the same trained stimulus across different trials, on-diagonal values in correlation matrix) compared with between-item similarity (maintenance of different trained stimuli, off-diagonal correlations), as schematized in Figure 4B (left). In order to provide the most straightforward and interpretable analysis of item-level representations, we focused on trained items that were not part of learned sequences. Analyzing these six stimuli (for each participant) avoids the potential confound that items in temporal sequences may restructure and develop more integrated or differentiated representations over time (Sakai and Miyashita, 1991; Schapiro et al., 2012; Schlichting et al., 2015). In this item-level model, higher pattern strength values correspond to stronger representations of individual items in WM (and differentiation from the other trained stimuli). This represents a critical test for the prediction of greater item-level selectivity in the lPFC with learning.

Pattern strength for the item-level model showed a significant increase over time in ventral mid-lateral lPFC (ventral mid-lateral: $r = 0.36$, $p = 0.005$ [FDR-corrected, $p = 0.036$];

Figure 4B, right; inflection point = 1.05 sessions) and not in other PFC or visual areas (all FDR-corrected, p values > 0.08). That is, patterns of WM delay activity for individual trained items became more robust (reliable across trials) and differentiated from other trained stimuli across learning. To ensure that this result was not driven by the logistic modeling approach, we confirmed that this increase in pattern strength was reliable using a linear model, which demonstrated a significant increase over sessions (Table S1). Finally, to test whether there were reliable differences in the similarity between each condition after learning began to unfold, we tested the difference between the on-diagonal (within-item) and off-diagonal (between-item) correlation values only from sessions occurring *after* the inflection point from the fitted model. This difference was significant in the ventral mid-lateral PFC, with individual items becoming more differentiated from other items (t = 2.11, p = 0.04). These analyses provide evidence for stronger item-specificity or differentiation of item-level representations in lPFC delay activity across the course of training.

We next asked whether WM representations of all items show evidence of neural differentiation over time, or whether this is specific to trained stimuli. If the item-specific representations in the lPFC are specific to trained stimuli, then activation patterns between trained stimuli should become less similar (as the items become more identifiable from each other), while those between novel stimuli should not reliably change. We operationalized this comparison with a category-level model, which tested for an interaction of a decrease in pattern similarity between trained stimuli (that were not part of sequences) relative to the change in similarity between novel stimuli (off-diagonal correlations), as schematized in Figure 4C (left). In this category-level model, higher pattern strength values correspond to a stronger differentiation between trained and novel stimuli across learning. There was a significant increase in pattern strength for the category-level model across sessions in multiple lPFC areas (dorsal rostral: r = 0.30, p = 0.015 [FDR-corrected, p = 0.030]; dorsal caudal: r = 0.34, p = 0.007 [FDR-corrected, p = 0.019]; ventral mid-lateral: r = 0.28, p = 0.025 [FDR-corrected, p = 0.039]; ventral caudal: r = 0.34, p = 0.007 [FDR-corrected, p = 0.019]) and the early visual cortex (early visual: r = 0.47, p < 0.001 [FDR-corrected p = 0.002]). The range of inflection points was between sessions 7.23–8.14 for these regions (Figure 4C, right). The category-level model also showed reliable differences in the similarity across conditions, with significantly lower between-item similarity for trained stimuli compared with novel stimuli after the inflection point for each region (dorsal rostral: t = −3.4, p = 0.002; dorsal caudal: t = −4.4, p = 0.0001; ventral mid-lateral: t = −3.2, p = 0.003; ventral caudal: t = −3.7, p = 0.001; early visual: t = −3.9, p = 0.0005). We also confirmed that linear modeling showed the increase in category-level pattern strength in the dorsal caudal PFC and early visual cortex, but not in the other ROIs (Table S1). These pattern similarity analyses reveal different representational information for trained and novel stimuli across learning, such that the distinction between trained (as compared to novel) stimuli becomes increasingly detectable over time.

Finally, we tested whether associations learned in a distinct task context may influence WM maintenance processes, even when they are not task-relevant. In parallel to the WM task, participants learned that a subset of trained stimuli formed high-probability temporal sequences in the SRT task (Figure S1). Based on classic studies of paired associate learning (Naya et al., 2001; Sakai and Miyashita, 1991) and multivariate representations that are altered by learning (Schapiro et al., 2012; Schlichting et al., 2015), we tested for shared representations across items in the same temporal sequence (higher similarity across items *within* the same sequence versus *between* sequences). Surprisingly, we found no increases in this representation over time. However, we did see a decrease in the early visual cortex, such that items in the same sequence became more distinct from each other over time (relative to between-sequence similarity; Figure S4).

To better understand any potential higher-level representations influenced by sequence learning, we lastly tested whether the organization of stimuli into temporal sequences in the SRT task may have resulted in a shared representation during WM. That is, a representation between stimuli belonging to *any* sequence (regardless of sequence identity) that is distinct from items that were not part of sequences (non-sequence items). This would reflect a categorical difference (or “boundary”) between trained items belonging to a sequence versus those without such associations. This coarse-level representation was operationalized with a sequence category model (Figure 5, left). The model compared changes in between-item similarity for items in sequences versus similarity for sequence-to-non-sequence items. Higher pattern strength values in the model correspond to stronger representations for the category of trained items in sequences from the SRT task, compared with the similarity of these items with other trained stimuli that were not part of temporal sequences.

Pattern strength for this sequence category model showed a significant increase across sessions in caudal lPFC regions (dorsal caudal: r = 0.41, p = 0.001 [FDR-corrected, p = 0.011]; ventral caudal: r = 0.36, p = 0.004 [FDR-corrected, p = 0.017]; Figure 5, right). The inflection points here were later than any other model (dorsal caudal: session 15.2, ventral caudal: session 10.5). As in the previous analyses, we also confirmed that the increase in pattern strength for the sequence category model in the caudal PFC ROIs was robust, using linear modeling (Table S1). Finally, there was a reliable difference in correlations in the conditions of the sequence category model; the similarity between-sequence stimuli was significantly higher than the mean similarity between-sequence and non-sequence stimuli after the inflection point (dorsal caudal: t = 4.28, p = 0.009; ventral caudal: t = 2.93, p = 0.008). Therefore, the sequence category representations were likely driven by higher similarity between trained stimuli that were part of sequences. Across these analyses that consider associations in the SRT task, stimuli across any learned sequence became more similar to each other over training, relative to stimuli not in sequences, specifically in caudal lPFC regions. These results suggest that learned associations from LTM are reflected in WM delay activity, even when those associations are irrelevant to WM task goals.

DISCUSSION

Here, we examined how long-term learning influences lPFC neural representations for WM. Over 3 months, we extensively trained three human participants on a WM task and a sequence learning task, which both employed a unique set of complex fractal stimuli. We sampled fMRI activity and behavioral performance repeatedly across learning and found that the distribution and selectivity of lPFC WM delay activity changed with training: more cortical territory was recruited during WM maintenance with learning, and these activity changes coincided with increases in stimulus representations in multivariate patterns (Figure

6). Associations between stimuli learned in another task context, although task-irrelevant for WM, also shaped neural representations in the lPFC during WM maintenance. In sum, long-term learning changed the distribution and representational structure of lPFC WM activity, indicating that the neural mechanisms for WM are influenced by prior experience.

lPFC: Representations or processes?

Early NHP recordings from the lPFC revealed neurons that respond to all phases of WM tasks: cue, delay, and response periods (Funahashi et al., 1990). Since then, neurons in the lPFC have been shown to encode stimulus representations (Funahashi et al., 1989; Murray et al., 2017), motor responses, task rules, and executive control signals (Rigotti et al., 2013; Vallentin et al., 2012; Wallis and Miller, 2003). In contrast, the human lPFC shows a relative absence of stimulus-specific representations during WM (D'Esposito and Postle, 2015; Harrison and Tong, 2009; Serences, 2016), while neuroimaging and lesion studies point to the lPFC as a source of cognitive control signals (Chatham et al., 2014; Gazzaley and Nobre, 2012; Szczepanski and Knight, 2014). Thus, the role of lPFC function during WM has been unclear across studies. However, NHP and human studies are characterized by stark differences in training regimes before neural recordings take place (Berger et al., 2018; Birman and Gardner, 2016; Sarma et al., 2016). Therefore, we reasoned that differences in task and stimulus experience may underlie the discrepant conclusions about lPFC function.

To directly test the influence of training on WM and lPFC function, we scanned participants across repeated WM task and stimulus exposure. As training progressed, we found that stimulus-specific information was increasingly represented in human lPFC delay activity, akin to patterns more commonly found in NHP studies. In human studies, WM content is typically more difficult to detect in the lPFC relative to visual areas (Bhandari et al., 2018; D'Esposito and Postle, 2015; Serences, 2016). However, a few prior studies also detect stimulus representations in the human lPFC, for instance, for visual orientations in retinotopically organized areas (Christophel et al., 2012; Ester et al., 2015) or object category (but not feature) information (Lee et al., 2013). Therefore, it may be that orientation stimuli or object categories are supported by a distributed, large-scale organization that enables information to be de-coded at the coarse level of fMRI, even without extensive training. Here, we show that individual representations for visually similar, complex images become increasingly detectable in the human lPFC, but not visual areas, through long-term learning. One possible explanation for this training-related change is that the WM information only became represented in the lPFC with learning. Another possibility is that there were preexisting WM representations, below the level of fMRI resolution, but learning altered the representational structure into a more detectable state (i.e., present in large-scale patterns). Likewise, in areas that show no detectable item-level or category representations here, WM representations may exist at finer levels of measurement granularity (at sub-voxel resolution, or in finer grained patterns that are absent across entire ROIs), or they may have been too noisy to yield reliable trends over time. Thus, null findings here should be interpreted with caution. Nonetheless, our results suggest that the debate over the role of the lPFC in WM may hinge on training. That is, delay period signals reflecting general WM maintenance (processes) are present in fMRI activity without extensive training, while responses to individual stimuli (representations) emerge in the lPFC after long-term learning.

Implications for models of lPFC functional organization

The lPFC is organized in a macroscale gradient along the rostral-caudal axis, both functionally (Badre and Nee, 2018; Koechlin et al., 2003) and anatomically (Goulas et al., 2014; Miller et al., 2021; Wagstyl et al., 2020). More abstract representations are generally encoded more rostrally along the lPFC (Badre and D’Esposito, 2009), with middle frontal areas posited to sit “atop” the hierarchy and provide top-down control during complex cognitive tasks (Badre and Nee, 2018; Duverne and Koechlin, 2017; Ito et al., 2017). Here, our data also support a rostral-caudal WM organization along the lPFC: after training, stimulus-specific representations emerged in the mid-lateral lPFC and categorical representations in caudal lPFC areas (Figure 6). The timeline of learning also reflected this abstraction, with progressively later inflection points for more categorical representations. Although stimulus categories are often more abstract than individual stimuli—and might therefore be expected to engage more rostral regions—the “categorical” model here may instead capture associations with motor planning processes for item sequences learned in the SRT task. This caudal lPFC sequence-level representation is also consistent with NHP studies finding categorical task and rule representations in homologous premotor areas (Muhammad et al., 2006; Vallentin et al., 2012; Wallis and Miller, 2003). Our results suggest that distinct levels of representation for learned stimuli during WM may be scaffolded onto an existing rostro-caudal lPFC functional organization.

Here, we show stimulus-specific activity patterns that are not typically detected in the human lPFC but are more akin to observations from NHP studies. However, the exact areas of functional homology are often observed to be anatomically distinct across the NHP and human lPFC. For example, lesions of caudal precentral areas in the human lPFC cause deficits in spatial WM that mirror the effects in NHPs of damage to more anterior lPFC areas (Mackey et al., 2016). Here, WM stimulus patterns emerge in micro-anatomically similar areas to where NHP recordings detect WM stimulus information (e.g., areas 9/46d, 9/46v; Petrides, 2005). Long-term learning drives mid-lateral lPFC regions—that are most often described as a “controller” of task activity in humans—to represent stimulus-specific WM information. This suggests the intriguing possibility that WM storage and top-down control signals can occur in the same location, depending on learning and task demands. Future work using longitudinal paradigms in NHP studies might also clarify the importance of training, spatial scale, and species differences on WM maintenance processes (Badre et al., 2015; Milham et al., 2018; Song et al., 2021).

PFC plasticity

The lPFC is critical for flexible cognition. Multiple theories consider the lPFC to have high plasticity, with activity patterns and representations that change based on task demands (Duncan, 2001; Miller and Cohen, 2001; Woolgar et al., 2011). However, these patterns of adaptation have not been systematically tracked over time in the human lPFC. Some human neuroimaging studies have employed forms of WM training as a route to improve WM and cognition more broadly, but the direction of change has been inconsistent. Early studies found that activation increases in the frontal and parietal cortex after WM training (Klingberg, 2010; Olesen et al., 2004), but recent aggregations of WM training studies roughly show that activation decreases for studies with shorter training times (~minutes-hours) and increases for longer training (~days-weeks) (Buschkuehl et al., 2012, 2014). These studies only sparsely sample neuroimaging data and have thus been unable to track learning across time or to examine the effects of stimulus experience and context on the neural mechanisms of WM maintenance. Here, we densely sampled neuroimaging and behavior across training, and showed progressive increases in both lPFC activity as well as multiple levels of stimulus representational information.

Recent NHP electrophysiology studies have observed changes after training in the selectivity and magnitude of both single-unit and population spiking during WM (Dang et al., 2021; Meyer et al., 2011; Tang et al., 2019; Tang et al., 2022). Long-term representations of learned stimulus categories are also detectable in the frontal and temporal cortex using NHP fMRI (Ghazizadeh et al., 2018), which may bridge between single-unit and human neuroimaging data. Here, we attempted to approximate the task difficulty and timeline of learning in NHP studies, and we chose a stimulus set of complex, novel fractals with which participants had no prior experience. The present data shows learning-related changes in the human lPFC that may parallel changes in NHP activity patterns and representations after training (Tang et al., 2019; Riley et al., 2018). However, it is difficult to fully bridge across discrepancies in measurement techniques and species: although BOLD fMRI signals can correlate with representations detected via multi-unit activity and high-gamma local field potential (LFP) (Klink et al., 2021; Manea et al., 2022), there is a complicated relationship between spiking activity, LFP signals, and BOLD measurements (Mukamel et al., 2005; Nir et al., 2007; Shi et al., 2017). Ultimately, paradigms using identical tasks, training timelines, and stimuli will be needed to compare the effects of learning on WM neural data between NHPs and humans.

Neurophysiological mechanisms behind activity changes with training, here and in previous studies, are difficult to ascertain. Changes in dopaminergic signaling and receptor sensitivity, along with correlated firing across neurons, may all play a mechanistic role in activation and selectivity increases of lPFC neuronal populations with training (Constantinidis and Klingberg, 2016; Riley et al., 2018; Vijayraghavan et al., 2007). These previous effects have been greatest in mid- and anterior dorsal areas of the lPFC, mirroring the organization of emerging stimulus-selective activity patterns that we observed here in the mid-lateral lPFC. This lPFC plasticity likely arises from several factors that give the region a high propensity for flexible representations: long-range anatomical connections (Chaudhuri et al., 2015; Wang et al., 2021), status as a hub between cortical networks (Bertolero et al., 2018; Fornito et al., 2019), and a late anatomical development (Garcia et al., 2018; García-Cabezas et al., 2019). Complementing the literature on flexible lPFC activity patterns based on task demands, here we show lPFC plasticity from experience and learning across months.

Influence of LTM on WM

According to foundational theories, WM and LTM are thought to rely on both different brain areas and neuronal mechanisms (Squire and Zola-Morgan, 1991; Warrington and Shallice, 1969; Wickelgren, 1996). Thus, the neural circuitry supporting WM has most often been studied without considering longer-term learning and memory effects. However, when WM behavior has been considered in relation to stimulus experience, better WM is observed for familiar, complex stimuli such as Pokémon (Xie and Zhang, 2017), meaningful human faces (Asp et al., 2021; Jackson and Raymond, 2008), and trained geometric shapes (Blalock, 2015). Our findings suggest that these experience-dependent WM behavioral changes are underpinned by malleability of the cortical representations that support WM across learning.

In addition to item-specific patterns, we found that shared WM representations develop for stimuli that were part of temporal sequences in the SRT task, consistent with a "categorical" representation grouping items based on their properties within learned knowledge structures. LTM consolidation is thought to promote the extraction of common features across experiences (Binder and Desai, 2011; Eichenbaum, 2017; McClelland et al., 1995; Winocur and Moscovitch, 2011); thus, it is likely that the shared categorical structure emerged as a function of memory consolidation, facilitated by repeated exposure to sequences over time (Antony et al., 2017). The later inflection points of the sequence category models (Figure 5) compared with the item-level model (Figure 4B) are consistent with gradual learning and consolidation facilitating the extraction of sequence information over time, which then influenced WM activity patterns. Item-level versus categorical representations also emerged in different areas of the lPFC, suggesting that the activity changes induced by long-term learning obeyed the functional axes of lPFC organization (Figure 6). Altogether, the results indicate not only that LTM can share representational formats with WM (Beukers et al., 2021; Lewis-Peacock and Norman, 2014; Nee and Jonides, 2011; Oberauer, 2009) but also that long-term learning *changes* how information is represented in WM, even when learned associations are not behaviorally relevant for WM.

Design considerations and caveats

In this study, we gained insight into how WM representations change with learning by using a longitudinal, dense sampling design within a targeted group. However, the limited sampling of few individuals results in findings that do not necessarily generalize to the population of healthy adults. Instead, we report detailed learning effects over months within a specific sample. Therefore, instead of examining variance between participants, we combined the data for participants, using a fixed-effects approach to make the most reliable and strongest inferences in the current sample (rather than the general population, see Fries and Maris, 2022; Vezoli et al., 2021). Similar designs involving a small group of individuals ("deep imaging") have been used to show changes in human brain activity and functional organization (Gordon et al., 2017; Naselaris et al., 2021; Newbold et al., 2020). These play a key role in modern neuroscience, placing emphasis on high data quality and within-participant power instead of sampling more individuals to achieve similar levels of statistical power (Gordon et al., 2017; Gratton and Braga, 2021; Gratton et al., 2022; Kragel et al., 2021; Naselaris et al., 2021; Newbold et al., 2020; Popham et al., 2021). We apply this "deep imaging" approach to tackle discrepancies between WM studies.

Future considerations

Here, we show that human lPFC activity patterns gradually change over long-term learning, suggesting that the role of the lPFC in WM may shift as stimuli become well-learned and embedded in associative structures. These findings highlight important considerations for conducting and interpreting investigations into WM function. If the neural circuitry for WM is shaped by prior experience, drastically different conclusions can be reached, depending on when brain recordings take place relative to training. The timeline of learning is especially important to consider because the lPFC ensembles demonstrate remarkable flexibility in activity patterns (e.g., magnitude, timing, and dimensionality) based on behavioral demands (Dang et al., 2021; Miller and Cohen, 2001; Miller and Fusi, 2013; Stokes et al., 2013; Wasmuht et al., 2018). By implementing a protracted training and recording regime in humans, our data show that long-term learning sculpts neural representations during WM. These data offer a potential bridge between seemingly incompatible accounts from NHP electrophysiology and human fMRI studies. Moving forward, an accurate understanding of PFC and WM functioning should consider training effects, species differences, and how LTM may be involved.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jacob Miller (j.a.miller@yale.edu).

Materials availability—This study did not generate new unique reagents.

Data and code availability—All neuroimaging data will be openly available in the Brain Imaging Data Structure format (Gorgolewski et al., 2016; https://bids.neuroimaging.io/) on OpenNeuro: https://openneuro.org/datasets/ds003659 (Markiewicz et al., 2021). Analysis and processing code to reproduce the present results, along with the stimuli, presentation code, and behavioral data may be found on Open Science Framework (OSF): https://doi.org/10.17605/OSF.IO/CKYBW.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human participants—The three study participants were all healthy, adult volunteers. Because of the large amount of MRI data collected and intensive nature of the behavioral training involved, all participants were members of the research team who completed the study over the same time period. While participants had limited experience performing WM and cognitive tasks before, this experience is much smaller relative to the study timeline, and they had no prior experience with this particular task or stimulus set. One participant was a 34-year-old female (sub-001), one was a 25-year-old male (sub-002), and one was a 37-year-old female (sub-003). The University of California, Berkeley Committee for the Protection of Human Subjects (CPHS) approved the study protocol and no participants reported any contraindications for MRI.

METHOD DETAILS

Study design and stimuli—The study was designed to investigate WM behavior and neural representations across a large amount of training on a specific set of stimuli and tasks. To accomplish this, we assigned each participant a unique set of 18 fractal images as their set of trained stimuli. Each image was an algorithmically-generated fractal consisting of multiple colors, and the 18 images for each participant were balanced according to the primary color group of each image (determined using a k-means clustering algorithm on each fractal image in the *sklearn* Python package: https://scikit-learn.org/). These fractals were chosen because they are visually complex, approximately uniform in size, cannot be easily verbalized, have no pre-existing meaning, and similar stimuli have been used in NHP electrophysiology studies of the neural basis of learning (Ghazizadeh et al., 2018; Kim et al., 2015; Sakai and Miyashita, 1991). Because the study participants were also on the research team, we avoided participants gaining any foreknowledge of their training set by generating thousands of initial images and randomly selecting each training set from among these images. Thus, each participants' first exposure to their training set occurred during the first scanning session. The unique 18 stimuli for each participant were then used for all of the following fMRI and behavioral training sessions, with additional novel stimuli randomly selected each session from the broader set of fractals. Of the 18 fractal stimuli in each participant's training set, 12 were randomly assigned to be part of four sequences in the SRT task, with each sequence consisting of three fractals and an object image. The sequences were learned over time as part of a serial reaction time (SRT) task. Although sequences were not explicitly instructed, all participants had knowledge of the sequence manipulation; thus, reductions in response time in this task likely reflect both explicit and implicit learning. All tasks were programmed using *Psychtoolbox* functions (Brainard, 1997; http://psychtoolbox.org/) in Matlab www.mathworks.com/), and stimuli were presented on a plain white background [RGB = 255,255,255].

Longitudinal training—Across the course of 15 weeks, each participant underwent 24-25 total sessions of fMRI scanning. In the present work, we analyze the first 17 of these fMRI sessions (*Phase 1*) for each participant which took place over ~3 months (13 weeks) of training. In a second study phase (*Phase 2*) of ~3 additional weeks, more fractal stimuli were added into the training set (Figure 1C), but the results from this phase of the experiment are not reported here. Over the first week, four scans were conducted to ensure that the initial exposure to the tasks and stimuli would be highly sampled. fMRI scanning during subsequent weeks occurred at a rate of approximately 1-2x per week (depending on participant and scanner availability). Before the study began, participants completed one block (24 trials) of WM task practice with pilot stimuli that never appeared in the main experiment.

To facilitate learning, at-home behavioral training was implemented multiple times per week across the course of the study (Figure 1C), where participants completed versions of the WM and sequence learning tasks on home laptop testing setups. Most sessions were completed at the same location for each subject, with a small number completed elsewhere (when traveling, for example). The at-home WM task training data can be found on Open Science Framework.

Working memory task—Participants completed a three-alternative forced choice delayed recognition task in each scanning and at-home WM training session (Figure 1A). Stimuli included the 18 fractals from the participant's training set, along with 6 novel fractal images, which were randomly selected for each session. On each trial, a single WM sample stimulus (600 x 600 pixels) was presented in the center of a screen for a 0.5 s encoding period. A fixation cross was then presented for a jittered delay period of 4, 8, or 12 s, with the goal of facilitating WM maintenance processes. A probe display then appeared for a response window of 2 s. The probe display comprised three occluded sections of fractal images ($^1/_6$ area of each image) at an equal distance from the center of the screen. Each probe image was masked within a gaussian window of FWHM at ~$^1/_6$ the image size. Participants responded via one of three button presses to indicate which probe image segment matched the stimulus from the beginning of the trial. A fourth button option could be used to indicate a guessing response of "I don't know". However, this option was rarely chosen (1.6% of total trials), so we could not examine meaningful changes in this response option across sessions. A sample-matching fractal image was always present in the probe display. One of the other probe stimuli was always a novel (untrained) fractal image randomly selected from the same color group as the sample fractal image. The third probe image was either a novel fractal (50% of trials) or a lure from the set of trained fractal images (50% of trials). The masked section of the fractal images was in the same location for each probe image and randomly chosen from nine different areas on each trial, and the probe position was counterbalanced across trials within a block (Figure 1A). After each trial, there was a jittered intertrial interval (ITI) sampled from an exponential distribution (mean = 4 s, range = 1 - 9 s).

In the scanning sessions, participants completed four blocks (scanner runs) of 24 trials, with each trained and novel fractal image presented as the WM sample stimulus once per block, in random order. Each delay length occurred in random order and equally often within a block. For the at-home WM training sessions, participants completed two blocks of 24 trials (Figure 1C). The in-scanner display was a back-projected 24 in. screen (1024 x 768) for an approximate ~47 cm viewing distance, while for at-home training sessions participants used laptop screens of sizes 13.3 in. (1440 x 900) [sub-001], 13.3 in. (2560 x 1600) [sub-002], and 12.5 in. (1920 x 1080) [sub-003].

Serial reaction time task—In addition to the WM task, participants completed a serial reaction time (SRT) task before the WM task in each scanning session and during at-home training sessions. This task served to repeatedly expose participants to statistical regularities amongst the trained stimuli, in the form of temporal stimulus sequences. During this task, participants made button presses in response to each stimulus. The stimulus set consisted of the same 18 fractal stimuli shown in the WM task as well as six objects (three animals and three tools) for a total of 24 stimuli. The SRT task consisted of two phases: an initial phase in which stimulus-response mappings were learned), followed by a second phase during which stimulus sequences were present.

The first section of SRT task was implemented in the first two sessions of the study (one fMRI session followed by an at-home behavioral session) during which participants were trained to criterion to associate each of the stimuli with one of four button press responses. Participants were first exposed to their stimulus set during their first scanning session.

During every block, each of the 24 stimuli were shown once in a randomized order, with no explicit sequence information present (during the first two sessions). Each stimulus was presented on the screen for 2.3 seconds (followed by a blank screen of 0.7 s between stimuli) with four response options shown as black squares below the stimulus (corresponding to the middle finger of the left hand, ring finger of the left hand, ring finger of the right hand, and middle finger of the right hand). During the first two blocks of the first scanning session, the correct response was highlighted (square corresponding to the response was shown in red instead of black) to allow participants to view the correct response and facilitate learning. Thereafter, participants completed 10 more blocks during which the correct response was not shown but feedback was provided (when a correct response was made the square turned blue and incorrect responses were indicated by the selected option turning red with feedback lasting for 200 ms). After the first scanning session, participants performed an at-home session to ensure the learning of stimulus-response mappings. Participants completed a minimum of five blocks of the task, and continued until a criterion of 80% accuracy at the item-level was reached (>=80% of correct first responses for all stimuli across all blocks; 7 - 15 blocks of training were required to reach criterion). The stimulus-response mappings remained constant throughout the study.

After the completion of training to criterion, temporal sequences of stimuli were embedded in the SRT task, beginning in the second fMRI session. Of the 24 trained stimuli (18 fractals and six objects), 16 stimuli were assigned to form four distinct sequences, with each sequence containing three fractals followed by an object (Figure 1B). As in the initial section of this task, each stimulus was shown once during each block (set of 24 trials) and the four response options were indicated below the stimulus as four black squares. Participants were instructed to press the appropriate button for each stimulus. Each stimulus was shown for 1.95 s (fMRI sessions) or 1.8 s (behavioral sessions) followed by a blank screen for 400 ms. Sequences were presented in a probabilistic manner, such that three of the four sequences were presented in an intact fashion in each block and each sequence was intact on 75% of blocks in each session (i.e. in 12/16 blocks during fMRI sessions). In each block, the order of the presentation of stimuli was randomized with the exception of the presentation of the three intact sequences. Stimuli from the non-intact sequence (one sequence per block) were presented in a random order with the stipulation that at least two stimuli separated the non-intact sequence stimuli. Feedback was provided throughout the experiment as described above in the training to criterion phase. The fMRI sessions contained 18 blocks of the SRT task and the at-home behavioral sessions consisted of 26 blocks. Stimuli were presented in a randomized order (no sequence information was present) during the first two blocks of each session which served to acclimate participants to the task.

Object-selective functional localizer task—Functional localizer scans were collected during two separate fMRI sessions for each participant, which occurred after sessions 1 and 5 for sub-001, sessions 1 and 15 for sub-002, and sessions 5 and 14 for sub-003. Participants performed a one-back task while viewing blocks of animals, tools, objects, faces, scenes, and scrambled images. All images were presented on phase scrambled backgrounds. Each block lasted for 16 s and contained 20 stimuli per block (300 ms stimulus presentation followed by a blank 500 ms inter-stimulus interval). Two stimuli were repeated in each block and participants were instructed to respond to stimulus repetitions via button press. Each scan (three scans per session) contained four blocks of each stimulus class, which were interleaved with five blocks of passive fixation.

fMRI acquisition—All neuroimaging data were collected on a 3 Tesla Siemens MRI scanner at the UC Berkeley Henry H. Wheeler Jr. Brain Imaging Center (BIC). Whole-brain Blood Oxygen Level-Dependent (BOLD) fMRI (T_2*-weighted) scans were acquired with a 32-channel RF head coil using a 2x accelerated multiband echo-planar imaging (EPI) sequence [repetition time (TR) = 2 s, echo time = 30.2 ms, flip angle (FA) = 80°, 2.5 mm isotropic voxels, 52 slices, matrix size = 84 x 84]. Anatomical MRI scans were collected at two timepoints across the study and registered and averaged together before further preprocessing. Each T_1-weighted anatomical MRI was collected with a 32-channel head coil using an MPRAGE gradient-echo sequence [repetition time (TR) = 2.3 s, echo time = 3 ms, 1 mm isotropic voxels]. For each scan, participants wore custom-fitted headcases (caseforge.com) to facilitate a consistent imaging slice prescription across sessions and to minimize head motion during data acquisition (Power et al., 2019).

In each 2-hr scanning session, participants completed the following BOLD fMRI scans: (1) 9 min eyes-closed rest run, (2) three 9 min runs of a 1-back stimulus localizer, (3) three 6 min runs of the SRT task, (4) 9 min eyes-closed rest block, (5) 9-min stimulus localizer block, (6) four 6 min runs of the WM task. The present work focuses on the WM task. In the stimulus localizer scans, participants viewed trained images in isolation in order to characterize how neural representations change over time (results not reported here).

QUANTIFICATION AND STATISTICAL ANALYSIS

fMRI preprocessing—Preprocessing of the neuroimaging data was performed using fMRIPrep version 1.4.0 (Esteban et al., 2019), a Nipype (Gorgolewski et al., 2017) based tool. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using *N4BiasFieldCorrection* v2.1.0 (Tustison et al., 2010) and skull-stripped using *antsBrainExtraction.sh* v2.1.0 (using the OASIS template). Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (Zhang et al., 2001) (FSL v5.0.9).

Functional data was slice time corrected using 3dTshift from AFNI v16.2.07 (Cox, 1996) and motion corrected using mcflirt (Jenkinson et al., 2002) (FSL v5.0.9). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve and Fischl, 2009) with 9 degrees of freedom, using flirt (FSL). Motion correcting transformations and BOLD-to-T1w transformation were concatenated and applied in a single step using *antsApplyTransforms* (ANTs v2.1.0) using Lanczos interpolation. Many internal operations of FMRIPREP use Nilearn (Abraham et al., 2014), principally within the BOLD-processing workflow. For more details of the pipeline see https://fmriprep.readthedocs.io/en/latest/workflows.html. Finally, spatial smoothing was only performed in a 4mm FWHM kernel along the cortical surface (https://github.com/mwaskom/lyman/tree/v2.0.0) for the mean univariate activity analysis (Figure 2), while all other analyses used unsmoothed data.

Region-of-Interest (ROI) selection—To generate cortical surface reconstructions, the T_1-weighted anatomical MRIs were processed through the FreeSurfer (surfer.nmr.mgh.harvard.edu/) *recon-all* pipeline for gray and white matter segmentation (Dale et al., 1999; Fischl et al., 1999a). To construct the lPFC ROIs, we sampled a recent multimodal areal parcellation of the human cerebral cortex (Glasser et al., 2016) onto each participant's native anatomical surface via cortex-based alignment (Fischl et al., 1999b). We merged these smaller parcels on the surface into six different lPFC ROIs (combined bilaterally), with two splits along the rostral-caudal axis and one split along the dorsal-ventral axis (Figure 2B). The caudal lPFC ROIs fall along the precentral sulcus and gyrus, with the most rostral ROIs ending in frontopolar cortex around the anterior ends of the inferior and superior frontal sulci. The split between dorsal-ventral ROIs roughly falls along the posterior middle frontal sulci, analogous microstructurally to the principal sulcus of macaques (Miller et al., 2021; Petrides, 2019), and the ROIs are bounded dorsally by the superior frontal gyrus and ventrally by the inferior frontal gyrus. This lPFC division into six areas was designed to align with NHP electrophysiology studies recording from multiple frontal cortex regions (Riley et al., 2018).

We also constructed two visual ROIs in order to determine if effects were specific to lPFC or also generalized to lower and higher-order visual areas. An early visual cortex ROI combined visual cortical areas V1-V4 for each participant, defined from aligning a probabilistic visual region atlas (Wang et al., 2015) onto each subject's native cortical surface using cortex-based alignment (Figure 4A). A higher-order visual ROI for the lateral occipital complex (LOC) was defined from a separate category localizer scanning session [block-level general linear model (GLM) with a contrast of responses of objects > scrambled objects] (see object-selective functional localizer task). Voxel responses were thresholded at p <.0001 and the ROI was restricted to voxels reaching this statistical threshold on the lateral surface of the occipital cortex and the posterior portion of the fusiform gyrus (Schwarzlose et al., 2008).

Mean WM delay activity across training—We constructed a separate event-related GLM in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) for each participant and session in order to compare activity levels across training. Separate boxcar regressors were constructed for the encoding (0.5 s), delay (4, 8, or 12 s), and probe (2 s) periods of the WM task, and all regressors were convolved with a standard double-gamma hemodynamic response function (HRF). Separate task event regressors were created for trained and novel fractals. For the session-level GLMs, all fourWM task runs in each session were concatenated with the *spm_fmri_concatenate* function. Six rigid-body motion parameters were included as nuisance regressors, along with high-pass filtering (HPF) of 128s to capture low-frequency trends as implemented in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). Voxelwise beta-coefficient and t-statistic maps were then calculated for WM delay (delay > fixation) periods, selecting regressors for trials across all three delay lengths. We analyzed changes in mean WM delay activity (beta coefficients) over learning with logistic models using mean activity in each ROI as the outcome variable and session number as the predictor (see statistical methods). These analyses were performed in two broad groups of voxels: (1) for the mean activity of voxels within the peak activation for each ROI (thresholding the maps for each participant and session at $t > 2.5$) and (2) for the mean activity of all voxels in each ROI (without any thresholding).

For each ROI, we also calculated the time course of activation across the encoding, delay, and probe period for trials with 12 s delay periods (Figure S6). Activation was plotted at each TR within the trial and normalized relative to a baseline of the mean signal during the ITI period.

Voxelwise regression analysis (recruitment of voxels across training)—To ask whether voxels showed changes in activity across training, we performed voxelwise logistic modeling on the *beta*-coefficient values from the above GLMs (*Mean WM delay activity across training*) across sessions (Figures 2A and 3). Separate voxelwise models were run on WM encoding and delay period activation to characterize changes in each phase of the WM task separately. For each participant and lPFC region, a logistic function was fit to model changes in activation across sessions. Goodness-of-fit of this model was assessed via the correlation between actual and predicted data (using 6-fold cross-validation, see statistical methods). Here, positive values indicate an increase in activity across sessions and negative values indicate a decrease in activity across sessions. After thresholding the voxelwise *r*-value maps ($p < 0.05$, see Figure 3C for maps for each participant), we then calculated the proportion of voxels in each ROI showing an increase or decrease in activity across sessions and averaged this value across participants. This generated a measure of how many voxels in an ROI change their activity over time, without requiring overlap of the specific voxels showing changes across participants. To determine if the proportion of voxels showing an increase or decrease in activity across sessions was different than chance ($p < 0.05 / 2 = 2.5\%$ false-alarm rate for increases or decreases), we constructed permuted null distributions of the proportion of increasing and decreasing voxels in each ROI. In each of 1,000 permutations, session number was randomly shuffled, the same logistic model fitting procedure was performed and regression of predicted activity onto activity across sessions was re-computed, and the proportion of voxels showing increases and decreases in activity (mean across participants) was stored to create null distributions. The true proportion of increasing and decreasing voxels across participants (dark lines in Figure 3B) was then compared to the null distributions obtained from permuting session labels to estimate *P*-values representing statistical significance of the proportion of voxels that changed over time.

Representational similarity analyses—In order to determine if WM delay activity showed representations for any specific fractal stimuli we obtained single-trial level voxel-wise activity maps by constructing separate least-squares-all (LSA) GLMs for each run, session, and participant (Mumford et al., 2012). Here, GLMs were constructed separately for each run in order to estimate pattern similarity between different runs, so that correlation measures aren't influenced by temporal autocorrelation within each functional scan (Mumford et al., 2014; Zeithamova et al., 2017). In each run-level GLM, the WM delay period events for each of the 24 unique stimuli were modeled as separate boxcar regressors (collapsed across delay lengths) and convolved with a HRF. The combined WM encoding (0.5 s) and probe (2 s) events were included as nuisance regressors, again split by trained and novel stimuli. Six rigid-body motion parameters were also included as nuisance regressors, along with high-pass filtering (HPF) of 128s to capture low-frequency trends. Voxelwise *beta*-coefficient maps from each trial were then used in the pattern similarity analyses.

Before estimating pattern similarity of the delay period activity in each ROI, we applied a multivariate noise decomposition algorithm to the single-trial WM delay period responses (Walther et al., 2016). *This process used the time-series of residuals from the LSA GLM for each run to account for noise variance within each ROI, resulting in activation patterns that are less biased by the noise structure. Then, for each session, we calculated between-run correlations (similarity) between the trials for all stimuli (18 trained, 6 novel fractals) across all six run-pair combinations. Correlation values were Fisher-z transformed, and then the mean of the between-run correlations (across run-pairs) generated a representational similarity or correlation matrix (Figure 5A). One total run across all sessions and participants was removed from calculation of between-run correlations because of a visual MR artifact (present in the raw functional data). To test for distinct representational structures in WM delay period patterns, we operationalized each of four potential representations as specific predictors of pattern similarity and then analyzed how the strength of each model changed across training. Each representational structure was coded using values of (*1, −1*) for specific stimulus pairs, and negative values were then re-coded such that the regressor values across all conditions summed to zero (i.e. equal weighting was given to positive and negative conditions). After constructing these matrices, they were then used as predictors of the similarity values (Fisher z-transformed pearson correlation), resulting in a model fit (“pattern strength”) for each representational structure. This procedure was performed for each session, participant, and ROI.

First, we constructed an *item-level* model for individual stimulus representations by comparing the on-diagonal correlations (between trials featuring the same stimulus) and off-diagonal correlations for the six trained stimuli not included in any of the learned sequences (Figure 4B). Second, we operationalized a category-level model by testing for an interaction in the off-diagonal correlations among all pairs of 18 trained (Figure 4C, dark blue) stimuli and the six novel (Figure 4C, light blue) stimuli within each session. Finally, we constructed a separate model to test for representations of stimulus sequences from the SRT task. A *sequence cafegory* model tested for an interaction in the similarity of stimuli between *different* sequences (Figure 5), compared to a baseline of the correlations between stimuli in sequences to the trained stimuli not in sequences. A final follow-up model directly tested within versus between-sequence stimulus correlations in a sequence identity model, with no differences found across conditions (Figure S4). For the analysis of off-diagonal correlations among trained stimuli in Figure 4A, we excluded the correlations between stimulus pairs within the same sequence from the SRT task.

To determine if there were changes in pattern similarity across training, we used fixed-effects logistic models with there *beta* values from the representational structure matrix regressor (“pattern strength”) as the outcome variable and session number as a predictor. For all models, ROIs with a significant change in the pattern strength across training (significant correlation between the predicted values from the logistic model and the actual data via cross-validation, see statistical methods) are bolded in Figures 4 and 5. We also included early visual and lateral occipital ROIs in the pattern similarity analyses to determine what representational changes are specific to the PFC versus early and higher-order sensory areas. Finally, for all RSA models we also test whether there were reliable differences in mean pattern similarity across conditions (different trial types) after learning began to unfold using only from sessions occurring after the inflection point from the fitted model. Note that this analysis is not circular as model fits were driven by variance in pattern strength values over time, and here we are examining reliable offsets (differences from 0) in pattern strength in a subset of sessions.

Statistical methods—All changes across training were analyzed using a fixed-effects, logistic model approach, implemented in Scipy's (https://www.scipy.org/) *optimize.curve_fit* function with four free parameters fit to the data: a slope (k), inflection point (x_T), and baseline (b) and asymptote (L) values (where x is session number and Y is the outcome variable, such as activation level or pattern strength).

$$Y = \frac{L}{1 + e^{k * (x_T - x)}} + b$$

The free parameters were fit within ranges of −10 to 17 for x_T (permitting values before the first session allows a function with relatively small changes in Y that occur early in learning and reach a plateau quickly, initial parameter value of median session number), 0 to 10 for k (initial value = 0.1), and dynamically adjusted from 2*min(Y) to 2*max(Y) for b (initial value = 0) and L (initial value = mean Y-value for 20% of latest data points). Code for the logistic fits is available here: https://github.com/arielletambini/logistic-model-fitting

The baseline (mean of session 1 and 2 value) was also subtracted before as part of fitting the model to track changes over time, akin to a linear fixed-effects analysis with a constant regressor for each subject. To avoid overfitting and best capture reliable trends across sessions, we implemented a 6-fold cross-validation scheme during model fitting. Unlike a typical cross-validation procedure in which each fold of the cross-validation is trained on contiguous data points, here the training and test data for each fold was evenly spaced across the course of learning (session number) in order to not systematically miss one portion of the longitudinal dataset in the training set, which could poorly capture learning-related variability. Specifically, the held out test sessions for each cross-validation fold were separated by six sessions. For example, in the first cross-validation fold, sessions 1, 7, and 13 were held out while all other sessions were used in model fitting, in the second cross-validation fold, sessions 2, 8, and 14 were held out, and so on. A total of six folds were used, with three sessions held out for 5 of 6 folds, and two sessions held out for one fold. For fixed-effects analyses in which models were fit on data across all participants (changes in pattern strength over time), the same session numbers were used as training and test data for all participants in each cross-validation fold, such that each fold was trained and tested on data from all participants. Changes over time were assessed by correlating the held out predicted logistic model values from each fold of cross-validation to the actual data values across all sessions (Figure 2A). We computed a one-sided test when assessing the correlation between predicted and actual values, such that correlation (r) values below 0, indicating a poor model fit, were set to 0. In order to easily interpret the direction of trends over time, we then coded the correlation value (always > or = 0) to correspond with the direction of changes across sessions. That is, we multiplied the correlation by −1 for decreases over time such that decreases over time show r < 0 and increases over time show r > 0. Model parameters were the median value across cross-validation folds. In the few cases where the logistic model failed to converge (Figure 2C, right: univariate changes in dorsal rostral PFC for all voxels), the model was run again without cross-validation to obtain fit parameters and report the model fit. Data were combined across all participants into one model as fixed-effects in order to assess changes over time, which is recommended for studies with a similar sample size, most often from non-human primate electrophysiology studies (Fries and Maris, 2022; Vezoli et al., 2021).

For each of the RSA analyses of pattern strength across learning (Figures 4 and 5), we also analyzed changes within ROIs that showed a significant logistic model fit with a different modeling approach (Table S1). Specifically, we implemented an ordinary-least-squares (OLS) regression with *pattern strength* as the outcome measure and *mean-centered session number* as the predictor, along with the square of session number (2nd order polynomial), as predictor variables. Data was combined across all subjects using a fixed-effects factor (constant for each subject).

For all statistical tests across ROIs, false-discovery rate (FDR) correction ($q = 0.05$) was used to adjust for multiple comparisons across the number of ROIs in each analysis (6 or 8). Neuroimaging files were loaded and operated on using the *Nilearn* package (https://nilearn.github.io/; Abraham et al., 2014). For all plots, error bands reflect bootstrapped 68% confidence intervals as implemented in the *Seaborn* package (Waskom, 2021).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) grants F32MH106280 to A.T., F32MH111204 to A.K., and RO1 MH63901 to M.D. We also thank Ian Ballard, Chris Gorgolewski, and Regina Lapate for their input on data analysis and Dan Lurie for assistance with data collection.

REFERENCES

Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, and Varoquaux G (2014). Machine learning for neuroimaging with scikit-learn. Front. Neuroinform 8, 14. 10.3389/fninf.2014.00014. [PubMed: 24600388]

Antony JW, Ferreira CS, Norman KA, and Wimber M (2017). Retrieval as a fast route to memory consolidation. Trends Cogn. Sci 21, 573–576. 10.1016/j.tics.2017.05.001. [PubMed: 28583416]

Asp IE, Störmer VS, and Brady TF (2021). Greater visual working memory capacity for visually matched stimuli when they are perceived as meaningful. J. Cogn. Neurosci 33, 902–918. 10.1162/jocn_a_01693. [PubMed: 34449847]

Badre D, and D'Esposito M (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? Nat. Rev. Neurosci 10, 659–669. 10.1038/nrn2667. [PubMed: 19672274]

Badre D, Frank MJ, and Moore CI (2015). Interactionist neuroscience. Neuron 88, 855–860. 10.1016/j.neuron.2015.10.021. [PubMed: 26637794]

Badre D, and Nee DE (2018). Frontal cortex and the hierarchical control of behavior. Trends Cogn. Sci 22, 170–188. 10.1016/j.tics.2017.11.005. [PubMed: 29229206]

Berger M, Calapai A, Stephan V, Niessing M, Burchardt L, Gail A, and Treue S (2018). Standardized automated training of rhesus monkeys for neuroscience research in their housing environment. J. Neurophysiol 119, 796–807. 10.1152/jn.00614.2017. [PubMed: 29142094]

Bertolero MA, Yeo BTT, Bassett DS, and D'Esposito M (2018). A mechanistic model of connector hubs, modularity and cognition. Nat. Hum. Behav 2, 765–777. 10.1038/s41562-018-0420-6. [PubMed: 30631825]

Beukers AO, Buschman TJ, Cohen JD, and Norman KA (2021). Is activity silent working memory simply episodic memory? Trends Cogn. Sci 25, 284–293. 10.1016/j.tics.2021.01.003. [PubMed: 33551266]

Bhandari A, Gagne C, and Badre D (2018). Just above chance: is it harder to decode information from prefrontal cortex hemodynamic activity patterns? J. Cogn. Neurosci 30, 1473–1498. 10.1162/jocn_a_01291. [PubMed: 29877764]

Binder JR, and Desai RH (2011). The neurobiology of semantic memory. Trends Cogn. Sci 15, 527–536. 10.1016/j.tics.2011.10.001. [PubMed: 22001867]

Birman D, and Gardner JL (2016). Parietal and prefrontal: categorical differences? Nat. Neurosci 19, 5–7. 10.1038/nn.4204. [PubMed: 26713741]

Blalock LD (2015). Stimulus familiarity improves consolidation of visual working memory representations. Atten. Percept. Psychophys 77, 1143–1158. 10.3758/s13414-014-0823-z. [PubMed: 25720758]

Brady TF, Störmer VS, and Alvarez GA (2016). Working memory is not fixed-capacity: more active storage capacity for real-world objects than for simple stimuli. Proc. Natl. Acad. Sci. USA 113, 7459–7464. 10.1073/pnas.1520027113. [PubMed: 27325767]

Brainard DH (1997). The Psychophysics Toolbox. Spat Vis 10, 433–436. [PubMed: 9176952]

Buschkuehl M, Hernandez-Garcia L, Jaeggi SM, Bernard JA, and Jonides J (2014). Neural effects of short-term training on working memory. Cogn. Affect. Behav. Neurosci 14, 147–160. 10.3758/s13415-013-0244-9. [PubMed: 24496717]

Buschkuehl M, Jaeggi SM, and Jonides J (2012). Neuronal effects following working memory training. Dev. Cogn. Neurosci 2, S167–S179. 10.1016/j.dcn.2011.10.001. [PubMed: 22682905]

Chatham CH, Frank MJ, and Badre D (2014). Corticostriatal output gating during selection from working memory. Neuron 81, 930–942. 10.1016/j.neuron.2014.01.002. [PubMed: 24559680]

Chaudhuri R, Knoblauch K, Gariel MA, Kennedy H, and Wang XJ (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. Neuron 88, 419–431. 10.1016/j.neuron.2015.09.008. [PubMed: 26439530]

Christophel TB, Hebart MN, and Haynes JD (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. J. Neurosci 32, 12983–12989. 10.1523/JNEUROSCI.0184-12.2012. [PubMed: 22993415]

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, and Haynes JD (2017). The distributed nature of working memory. Trends Cogn. Sci 21, 111–124. 10.1016/j.tics.2016.12.007. [PubMed: 28063661]

Constantinidis C, Funahashi S, Lee D, Murray JD, Qi X-L, Wang M, and Arnsten AFT (2018). Persistent spiking activity underlies working memory. J. Neurosci 38, 7020–7028. 10.1523/jneurosci.2486-17.2018. [PubMed: 30089641]

Constantinidis C, and Klingberg T (2016). The neuroscience of working memory capacity and training. Nat. Rev. Neurosci 17, 438–449. 10.1038/nrn.2016.43. [PubMed: 27225070]

Cox RW (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res 23, 162–173. 10.1006/cbmr.1996.0014.

Curtis CE, and Sprague TC (2021). Persistent activity during working memory from front to back. Front. Neural Circuits 15, 696060. [PubMed: 34366794]

D'Esposito M, and Postle BR (2015). The cognitive neuroscience of working memory. Annu. Rev. Psychol 66, 115–142. 10.1146/annurev-psych-010814-015031. [PubMed: 25251486]

Dale AM, Fischl B, and Sereno MI (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179–194. 10.1006/nimg.1998.0395. [PubMed: 9931268]

Dang W, Jaffe RJ, Qi X-L, and Constantinidis C (2021). Emergence of nonlinear mixed selectivity in prefrontal cortex after training. J. Neurosci 41, 7420–7434. 10.1523/JNEUROSCI.2814-20.2021. [PubMed: 34301827]

Duncan J (2001). An adaptive coding model of neural function in prefrontal cortex. Nat. Rev. Neurosci 2, 820–829. 10.1038/35097575. [PubMed: 11715058]

Duverne S, and Koechlin E (2017). Rewards and cognitive control in the human prefrontal cortex. Cereb. Cortex 27, 5024–5039. 10.1093/cercor/bhx210. [PubMed: 28922835]

Eichenbaum H (2017). Prefrontal-hippocampal interactions in episodic memory. Nat. Rev. Neurosci 18, 547–558. 10.1038/nrn.2017.74. [PubMed: 28655882]

Esteban O, Markiewicz C, Blair RW, Moodie C, Isik AI, Erramuzpe Aliaga A, Kent J, Goncalves M, DuPre E, Snyder M, et al. (2019). FMRIPrep: a robust preprocessing pipeline for functional MRI. bioRxiv. 10.1101/306951.

Ester EF, Sprague TC, and Serences JT (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. Neuron 87, 893–905. 10.1016/j.neuron.2015.07.013. [PubMed: 26257053]

Fischl B, Sereno MI, and Dale AM (1999a). Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207. 10.1006/nimg.1998.0396. [PubMed: 9931269]

Fischl B, Sereno MI, Tootell RBH, and Dale AM (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp 8, 272–284. [PubMed: 10619420]

Fornito A, Arnatkevičiūtė A, and Fulcher BD (2019). Bridging the gap between connectome and transcriptome. Trends Cogn. Sci 23, 34–50 10.1016/j.tics.2018.10.005. [PubMed: 30455082]

Fries P, and Maris E (2022). What to do if N is two? J. Cogn. Neurosci 34, 1114–1118. [PubMed: 35468209]

Fukuda K, and Woodman GF (2017). Visual working memory buffers information retrieved from visual long-term memory. Proc. Natl. Acad. Sci. USA 114, 5306–5311. 10.1073/pnas.1617874114. [PubMed: 28461479]

Funahashi S, Bruce CJ, and Goldman-Rakic PS (1989). Mnemonic encoding of visual space in the monkey’s dorsolateral prefrontal cortex. J. Neurophysiol 61, 331–349. [PubMed: 2918358]

Funahashi S, Bruce CJ, and Goldman-Rakic PS (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. J. Neurophysiol 63, 814–831. 10.1152/jn.1990.63.4.814. [PubMed: 2341879]

Fuster JM, and Alexander GE (1971). Neuron activity related to short-term memory. Science 173, 652–654. [PubMed: 4998337]

Garcia KE, Robinson EC, Alexopoulos D, Dierker DL, Glasser MF, Coalson TS, Ortinau CM, Rueckert D, Taber LA, Van Essen DC, et al. (2018). Dynamic patterns of cortical expansion during folding of the preterm human brain. Proc. Natl. Acad. Sci. USA 115, 3156–3161. 10.1073/pnas.1715451115. [PubMed: 29507201]

Garcéa-Cabezas MÁ, Zikopoulos B, and Barbas H (2019). The structural model: a theory linking connections, plasticity, pathology, development and evolution of the cerebral cortex. Brain Struct. Funct 224, 985–1008. 10.1007/s00429-019-01841-9. [PubMed: 30739157]

Gazzaley A, and Nobre AC (2012). Top-down modulation: bridging selective attention and working memory. Trends Cogn. Sci 16, 129–135 10.1016/j.tics.2011.11.014. [PubMed: 22209601]

Ghazizadeh A, Griggs W, Leopold DA, and Hikosaka O (2018). Temporal-prefrontal cortical network for discrimination of valuable objects in long-term memory. Proc. Natl. Acad. Sci. USA 115, E2135–E2144. 10.1073/pnas.1707695115. [PubMed: 29437980]

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature 536, 171–178. 10.1038/nature18933. [PubMed: 27437579]

Goldman-Rakic PS (1995). Cellular basis of working memory. Neuron 14, 477–485. [PubMed: 7695894]

Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, et al. (2017). Precision functional mapping of individual human brains. Neuron 95, 791–807.e7. 10.1016/j.neuron.2017.07.011. [PubMed: 28757305]

Gorgolewski KJ, Auer T, Calhoun VD, Cameron Craddock R, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data 3, 1–9. 10.1038/sdata.2016.44.

Gorgolewski KJ, Esteban O, Ellis DG, Notter MP, Ziegler E, Johnson H, Hamalainen C, Yvernault B, Burns C, Manhães-Savio A, et al. (2017). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1. Front Neuroinform. 5, 13.

Goulas A, Uylings HB, and Stiers P (2014). Mapping the hierarchical layout of the structural network of the macaque prefrontal cortex. Cereb. Cortex 24, 1178–1194. 10.1093/cercor/bhs399. [PubMed: 23258344]

Gratton C, and Braga RM (2021). Editorial overview: deep imaging of the individual brain: past, practice, and promise. Curr. Opin. Behav. Sci 40. iii–vi. 10.1016/j.cobeha.2021.06.011.

Gratton C, Nelson SM, and Gordon EM (2022). Brain-behavior correlations: two paths toward reliability. Neuron 110, 1446–1449. 10.1016/j.neuron.2022.04.018. [PubMed: 35512638]

Greve DN, and Fischl B (2009). Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72. 10.1016/j.neuroimage.2009.06.060. [PubMed: 19573611]

Harrison SA, and Tong F (2009). Decoding reveals the contents of visual working memory in early visual areas. Nature 458, 632–635. 10.1038/nature07832. [PubMed: 19225460]

Hoskin AN, Bornstein AM, Norman KA, and Cohen JD (2019). Refresh my memory: episodic memory reinstatements intrude on working memory maintenance. Cogn. Affect. Behav. Neurosci 19, 338–354. 10.3758/s13415-018-00674-z. [PubMed: 30515644]

Ito T, Kulkarni KR, Schultz DH, Mill RD, Chen RH, Solomyak LI, and Cole MW (2017). Cognitive task information is transferred between brain regions via resting-state network topology. Nat. Commun 8, 1027. 10.1038/s41467-017-01000-w. [PubMed: 29044112]

Jackson MC, and Raymond JE (2008). Familiarity enhances visual working memory for faces. J. Exp. Psychol. Hum. Percept. Perform 34, 556–568. 10.1037/0096-1523.34.3.556. [PubMed: 18505323]

Jenkinson M, Bannister P, Brady M, and Smith S (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841. 10.1016/s1053-8119(02)91132-8. [PubMed: 12377157]

Kim HF, Ghazizadeh A, and Hikosaka O (2015). Dopamine neurons encoding long-term memory of object value for habitual behavior. Cell 163, 1165–1175. 10.1016/j.cell.2015.10.063. [PubMed: 26590420]

Klingberg T (2010). Training and plasticity of working memory. Trends Cogn. Sci 14, 317–324. 10.1016/j.tics.2010.05.002. [PubMed: 20630350]

Klink PC, Chen X, Vanduffel W, and Roelfsema PR (2021). Population receptive fields in nonhuman primates from whole-brain fMRI and large-scale neurophysiology in visual cortex. eLife 10. 10.7554/eLife.67304.

Koechlin E, Ody C, and Kouneiher F (2003). The architecture of cognitive control in the human prefrontal cortex. Science 302, 1181–1185. 10.1126/science.1088545. [PubMed: 14615530]

Kragel PA, Han X, Kraynak TE, Gianaros PJ, and Wager TD (2021). Functional MRI can be highly reliable, but it depends on what you measure: a commentary on Elliott et al. (2020). Psychol. Sci 32, 622–626. 10.1177/0956797621989730. [PubMed: 33685310]

Lara AH, and Wallis JD (2015). The role of prefrontal cortex in working memory: a mini review. Front. Syst. Neurosci 9, 173. 10.3389/fnsys.2015.00173. [PubMed: 26733825]

LaRocque JJ, Lewis-Peacock JA, and Postle BR (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. Front. Hum. Neurosci 8, 5. [PubMed: 24478671]

Leavitt ML, Mendoza-Halliday D, and Martinez-Trujillo JC (2017). Sustained activity encoding working memories: not fully distributed. Trends Neurosci. 40, 328–346. 10.1016/j.tins.2017.04.004. [PubMed: 28515011]

Lee SH, Kravitz DJ, and Baker CI (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat. Neurosci 16, 997–999. 10.1038/nn.3452. [PubMed: 23817547]

Lewis-Peacock JA, and Norman KA (2014). Competition between items in working memory leads to forgetting. Nat. Commun 5, 5768. 10.1038/ncomms6768. [PubMed: 25519874]

Lorenc ES,and Sreenivasan KK(2021). Reframing the debate: the distributed systems view of working memory. Vis. Cogn 1–21. 10.1080/13506285.2021.1899091. [PubMed: 33574729]

Mackey WE, Devinsky O, Doyle WK, Meager MR, and Curtis CE (2016). Human dorsolateral prefrontal cortex is not necessary for spatial working memory. J. Neurosci 36, 2847–2856. 10.1523/JNEUROSCI.3618-15.2016. [PubMed: 26961941]

Manea AMG, Zilverstand A, Ugurbil K, Heilbronner SR, and Zimmermann J (2022). Intrinsic timescales as an organizational principle of neural processing across the whole rhesus macaque brain. eLife 11, e75540. 10.7554/eLife.75540. [PubMed: 35234612]

Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, Hardcastle N, Wexler J, Esteban O, Goncavles M, et al. (2021). The OpenNeuro resource for sharing of neuroscience data. eLife 10, e71774. 10.7554/eLife.71774. [PubMed: 34658334]

McClelland JL, McNaughton BL, and O'Reilly RC (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev 102, 419–457. 10.1037/0033-295X.102.3.419. [PubMed: 7624455]

Mendoza-Halliday D, Torres S, and Martinez-Trujillo JC (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. Nat. Neurosci 17, 1255–1262. 10.1038/nn.3785. [PubMed: 25108910]

Meyer T, Qi X-L, Stanford TR, and Constantinidis C (2011). Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks. J. Neurosci 31, 6266–6276. 10.1523/JNEUROSCI.6798-10.2011. [PubMed: 21525266]

Milham MP, Ai L, Koo B, Xu T, Amiez C, Balezeau F, Baxter MG, Blezer ELA, Brochier T, Chen A, et al. (2018). An open resource for non-human primate imaging. Neuron 100, 61–74.e2. 10.1016/j.neuron.2018.08.039. [PubMed: 30269990]

Miller EK, and Cohen JD (2001). An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci 24, 167–202. [PubMed: 11283309]

Miller EK, and Fusi S (2013). Limber neurons for a nimble mind. Neuron 78, 211–213. 10.1016/j.neuron.2013.04.007. [PubMed: 23622059]

Miller EK, Lundqvist M, and Bastos AM (2018). Working Memory 2.0. Neuron 100, 463–475. 10.1016/j.neuron.2018.09.023. [PubMed: 30359609]

Miller JA, Voorhies WI, Lurie DJ, D'Esposito M, and Weiner KS (2021). Overlooked tertiary sulci serve as a meso-scale link between microstructural and functional properties of human lateral prefrontal cortex. J. Neurosci 41, 2229–2244. 10.1523/JNEUROSCI.2362-20.2021. [PubMed: 33478989]

Muhammad R, Wallis JD, and Miller EK (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. J. Cogn. Neurosci 18, 974–989. 10.1162/jocn.2006.18.6.974. [PubMed: 16839304]

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, and Malach R (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. Science 309, 951–954. 10.1126/science.1110913. [PubMed: 16081741]

Mumford JA, Davis T, and Poldrack RA (2014). The impact ofstudy design on pattern estimation for single-trial multivariate pattern analysis. Neuroimage 103, 130–138. 10.1016/j.neuroimage.2014.09.026. [PubMed: 25241907]

Mumford JA, Turner BO, Ashby FG, and Poldrack RA (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59, 2636–2643. 10.1016/j.neuroimage.2011.08.076. [PubMed: 21924359]

Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, and Wang XJ (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. Proc. Natl. Acad. Sci. USA 114, 394–399. 10.1073/pnas.1619449114. [PubMed: 28028221]

Naselaris T, Allen E, and Kay K (2021). Extensive sampling for complete models of individual brains. Curr. Opin. Behav. Sci 40, 45–51. 10.1016/j.cobeha.2020.12.008.

Naya Y, Yoshida M, and Miyashita Y (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. Science 291, 661–664. 10.1126/science.291.5504.661. [PubMed: 11158679]

Nee DE, and Jonides J (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory: evidence for a 3-state model of memory. Neuroimage 54, 1540–1548. 10.1016/j.neuroimage.2010.09.002. [PubMed: 20832478]

Newbold DJ, Laumann TO, Hoyt CR, Hampton JM, Montez DF, Raut RV, Ortega M, Mitra A, Nielsen AN, Miller DB, et al. (2020). Plasticity and spontaneous activity pulses in disused human brain circuits. Neuron 107, 580–589.e6. 10.1016/j.neuron.2020.05.007. [PubMed: 32778224]

Nir Y, Fisch L, Mukamel R, Gelbard-Sagiv H, Arieli A, Fried I, and Malach R (2007). Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. Curr. Biol 17, 1275–1285. 10.1016/j.cub.2007.06.066. [PubMed: 17686438]

Oberauer K (2009). Design for a working memory. Psychol. Learn. Motiv 51, 45–100.

Olesen PJ, Westerberg H, and Klingberg T (2004). Increased prefrontal and parietal activity after training of working memory. Nat. Neurosci 7, 75–79. 10.1038/nn1165. [PubMed: 14699419]

Park SH, Russ BE, McMahon DBT, Koyano KW, Berman RA, and Leopold DA (2017). Functional subpopulations of neurons in a macaque face patch revealed by single-unit fMRI mapping. Neuron 95, 971–981.e5. 10.1016/j.neuron.2017.07.014. [PubMed: 28757306]

Petrides M (2005). Lateral prefrontal cortex: architectonic and functional organization. Philos. Trans. R. Soc. Lond. B Biol. Sci 360, 781–795. 10.1098/rstb.2005.1631. [PubMed: 15937012]

Petrides M (2019). Atlas of the Morphology of the Human Cerebral Cortex on the Average MNI Brain (Elsevier).

Popham SF, Huth AG, Bilenko NY, Deniz F, Gao JS, Nunez-Elizalde AO, and Gallant JL (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. Nat. Neurosci 24, 1628–1636. 10.1038/s41593-021-00921-6. [PubMed: 34711960]

Power JD, Silver BM, Silverman MR, Ajodan EL, Bos DJ, and Jones RM (2019). Customized head molds reduce motion during resting state fMRI scans. Neuroimage 189, 141–149. 10.1016/j.neuroimage.2019.01.016. [PubMed: 30639840]

Ranganath C, and Blumenfeld RS (2005). Doubts about double dissociations between short- and long-term memory. Trends Cogn. Sci 9, 374–380. 10.1016/j.tics.2005.06.009. [PubMed: 16002324]

Ranganath C, Johnson MK, and D’Esposito M (2003). Prefrontal activity associated with working memory and episodic long-term memory. Neuropsychologia 41, 378–389. [PubMed: 12457762]

Riggall AC, and Postle BR (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J. Neurosci 32, 12990–12998. 10.1523/jneurosci.1892-12.2012. [PubMed: 22993416]

Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, and Fusi S (2013).The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585–590. 10.1038/nature12160. [PubMed: 23685452]

Riley MR, Qi XL, Zhou X, and Constantinidis C (2018). Anterior-posterior gradient of plasticity in primate prefrontal cortex. Nat. Commun 9, 3790. 10.1038/s41467-018-06226-w. [PubMed: 30224705]

Romo R, Brody CD, Hernández A, and Lemus L (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. Nature 399, 470–473. 10.1038/20939. [PubMed: 10365959]

Sakai K, and Miyashita Y (1991). Neural organization for the long-term memory of paired associates. Nature 354, 152–155. 10.1038/354152a0. [PubMed: 1944594]

Sarma A, Masse NY, Wang X-J, and Freedman DJ (2016). Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. Nat. Neurosci 19, 143–149. 10.1038/nn.4168. [PubMed: 26595652]

Schapiro AC, Kustner LV, and Turk-Browne NB (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. Curr. Biol 22, 1622–1627. 10.1016/j.cub.2012.06.056. [PubMed: 22885059]

Schlichting ML, Mumford JA, and Preston AR (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun 6, 8151. 10.1038/ncomms9151. [PubMed: 26303198]

Schwarzlose RF, Swisher JD, Dang S, and Kanwisher N (2008). The distribution of category and location information across object-selective regions in human visual cortex. Proc. Natl. Acad. Sci. USA 105, 4447–4452. 10.1073/pnas.0800431105. [PubMed: 18326624]

Serences JT (2016). Neural mechanisms of information storage in visual short-term memory. Vision Res. 128, 53–67. 10.1016/j.visres.2016.09.010. [PubMed: 27668990]

Shi Z, Wu R, Yang P-F, Wang F, Wu T-L, Mishra A, Chen LM, and Gore JC (2017). High spatial correspondence at a columnar level between activation and resting state fMRI signals and local field potentials. Proc. Natl. Acad. Sci. USA 114, 5253–5258. 10.1073/pnas.1620520114. [PubMed: 28461461]

Song X, García-Saldivar P, Kindred N, Wang Y, Merchant H, Meguerditchian A, Yang Y, Stein EA, Bradberry CW, Ben Hamed S, et al. (2021). Strengths and challenges of longitudinal non-human primate neuroimaging. Neuroimage 236, 118009. 10.1016/j.neuroimage.2021.118009. [PubMed: 33794361]

Squire LR, and Zola-Morgan S (1991). The medial temporal lobe memory system. Science 253, 1380–1386. 10.1126/science.1896849. [PubMed: 1896849]

Sreenivasan KK, Curtis CE, and D'Esposito M (2014). Revisiting the role of persistent neural activity during working memory. Trends Cogn. Sci 18, 82–89. 10.1016/j.tics.2013.12.001. [PubMed: 24439529]

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, and Duncan J (2013). Dynamic coding for cognitive control in prefrontal cortex. Neuron 78, 364–375. 10.1016/j.neuron.2013.01.039. [PubMed: 23562541]

Supèr H, Spekreijse H, and Lamme VA (2001). A neural correlate of working memory in the monkey primary visual cortex. Science 293, 120–124. 10.1126/science.1060496. [PubMed: 11441187]

Szczepanski SM, and Knight RT (2014). Insights into human behavior from lesions to the prefrontal cortex. Neuron 83, 1002–1018. 10.1016/j.neuron.2014.08.011. [PubMed: 25175878]

Tang H, Qi XL, Riley MR, and Constantinidis C (2019). Working memory capacity is enhanced by distributed prefrontal activation and invariant temporal dynamics. Proc. Natl. Acad. Sci. USA 116, 7095–7100. [PubMed: 30877250]

Tang H, Riley MR, Singh B, Qi X-L, Blake DT, and Constantinidis C (2022). Prefrontal cortical plasticity during learning of cognitive tasks. Nat. Commun 13, 90. 10.1038/s41467-021-27695-6. [PubMed: 35013248]

Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, and Gee JC (2010). N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, 1310–1320. 10.1109/TMI.2010.2046908. [PubMed: 20378467]

Vallentin D, Bongard S, and Nieder A (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. J. Neurosci 32, 6621–6630. 10.1523/JNEUROSCI.5071-11.2012. [PubMed: 22573684]

Vezoli J, Vinck M, Bosman CA, Bastos AM, Lewis CM, Kennedy H, and Fries P (2021). Brain rhythms define distinct interaction networks with differential dependence on anatomy. Neuron 109, 3862–3878.e5. 10.1016/j.neuron.2021.09.052. [PubMed: 34672985]

Vijayraghavan S, Wang M, Birnbaum SG, Williams GV, and Arnsten AFT (2007). Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. Nat. Neurosci 10, 376–384. 10.1038/nn1846. [PubMed: 17277774]

Wagstyl K, Larocque S, Cucurull G, Lepage C, Cohen JP, Bludau S, Palomero-Gallagher N, Lewis LB, Funck T, Spitzer H, et al. (2020). BigBrain 3D atlas of cortical layers: cortical and laminar thickness gradients diverge in sensory and motor cortices. PLoS Biol. 18, e3000678. 10.1371/journal.pbio.3000678. [PubMed: 32243449]

Wallis JD, and Miller EK (2003). From rule to response: neuronal processes in the premotor and prefrontal cortex. J. Neurophysiol 90, 1790–1806. 10.1152/jn.00086.2003. [PubMed: 12736235]

Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, and Diedrichsen J (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage 137, 188–200. 10.1016/j.neuroimage.2015.12.012. [PubMed: 26707889]

Wang L, Mruczek RE, Arcaro MJ, and Kastner S (2015). Probabilistic maps of visual topography in human cortex. Cereb. Cortex 25, 3911–3931. 10.1093/cercor/bhu277. [PubMed: 25452571]

Wang Y, Royer J, Park B-Y, de Wael RV, Lariviere S, Tavakol S, Rodriguez-Cruces R, Paquola C, Hong S-J, Margulies D, et al. (2021). Long-range connections mirror and link microarchitectural and cognitive hierarchies in the human brain *2022*, bhac172.

Warrington EK, and Shallice T (1969). The selective impairment of auditory verbal short-term memory. Brain 92, 885–896. 10.1093/brain/92.4.885. [PubMed: 5364015]

Waskom M (2021). seaborn: statistical data visualization. J. Open Source Softw 6, 3021. 10.21105/joss.03021.

Wasmuht DF, Spaak E, Buschman TJ, Miller EK, and Stokes MG (2018). Intrinsic neuronal dynamics predict distinct functional roles during working memory. Nat. Commun 9, 3499. 10.1038/s41467-018-05961-4. [PubMed: 30158572]

Wickelgren WA (1996). Sparing of short-term memory in an amnesic patient: implications for strength theory of memory. Neurocase 2, 259as–298. 10.1093/neucas/2.4.259-as.

Winocur G, and Moscovitch M (2011). Memory transformation and systems consolidation. J. Int. Neuropsychol. Soc 17, 766–780. 10.1017/S1355617711000683. [PubMed: 21729403]

Woolgar A, Hampshire A, Thompson R, and Duncan J (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. J. Neurosci 31, 14592–14599. 10.1523/JNEUROSCI.2616-11.2011. [PubMed: 21994375]

Xie W, and Zhang W (2017). Familiarity increases the number of remembered Pokemon in visual short-term memory. Mem. Cognit 45, 677–689. 10.3758/s13421-016-0679-7.

Yonelinas AP (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. Behav. Brain Res 254, 34–44. 10.1016/j.bbr.2013.05.030. [PubMed: 23721964]

Zeithamova D, deAraujoSanchez MA, and Adke A (2017). Trial timing and pattern-information analyses offMRI data. Neuroimage 153,221–231. 10.1016/j.neuroimage.2017.04.025. [PubMed: 28411155]

Zhang Y, Brady M, and Smith S (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57. 10.1109/42.906424. [PubMed: 11293691]

Highlights

- Long-term training alters prefrontal cortex function during working memory
- Representations for individual stimuli in working memory emerge in prefrontal cortex
- Working memory representations are shaped by long-term associative learning
- Learning may reconcile debates on the role of prefrontal cortex in working memory

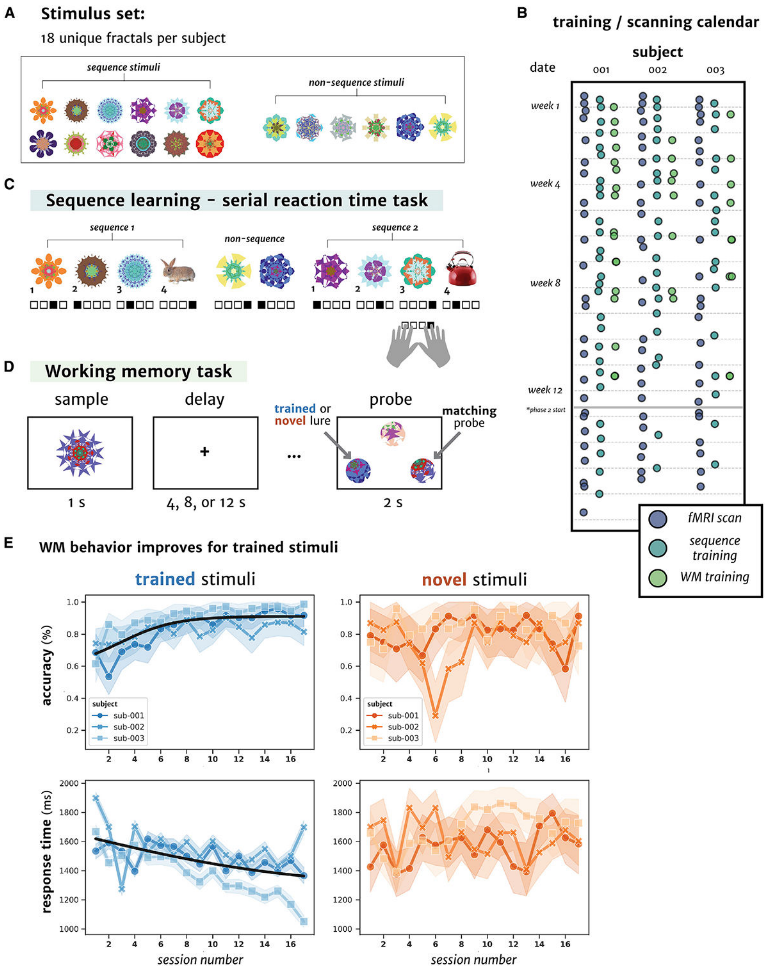

Figure 1. Longitudinal training across 3 months within individuals

(A) Example set of 18 unique fractal stimuli assigned to a single participant for the in-scanner and at-home behavioral tasks.

(B) Calendar of all of the MRI (purple) and at-home sessions (SRT, dark green; WM, light green) for each of the three participants over the 4 months of the study. During each MRI session, participants completed both the sequence learning and WM tasks. This study analyzes the first 17 sessions, as afterward new stimuli were added into the training set for each participant.

(C) The SRT task, in which each of the 18 trained stimuli was associated with one of the four button responses. Of the 18 trained stimuli, 12 were part of 4 sequences that occurred with high probability (75%) in the SRT task, and participants learned the sequences over time (Figure S1).

(D) The delayed three-alternative forced choice WM task, in which one fractal (trained or novel) was presented on each trial. After a jittered delay, participants indicated which occluded image matched the original sample.

(E) WM task accuracy (top) and response time (bottom) improved across training (sessions 1–17) for trials with one of the 18 trained stimuli (blue) but not for trials with novel fractal stimuli (orange). Accuracy error bars represent a bootstrapped 68% confidence interval (CI) across different blocks within each session (4 per participant per session), while for RT, error bars (bootstrapped 68% CI) are plotted across trials within each session for each participant.

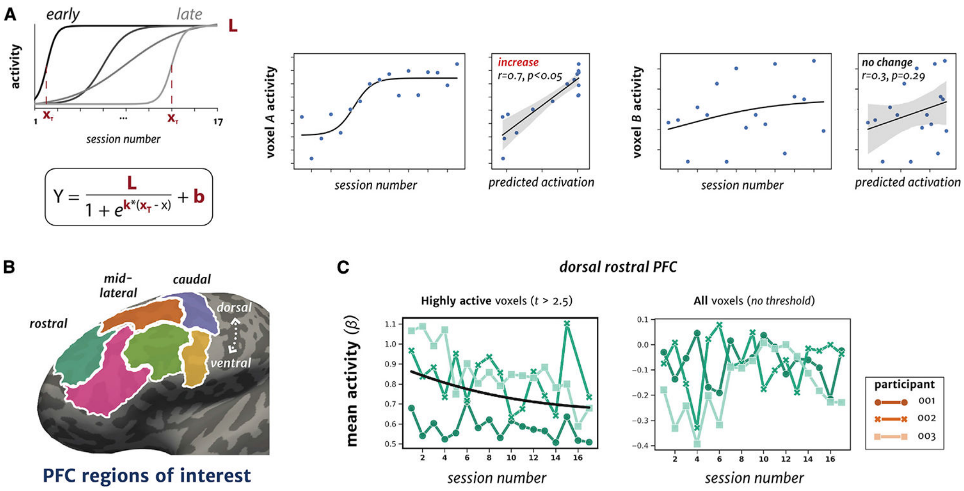

Figure 2. Mean WM delay activity changes in the PFC across the course of learning

(A) Schematic of the logistic modeling approach, in which four free model parameters were fit to the mean WM delay activity from each voxel across sessions (STAR Methods). Left inset, example WM delay activity profiles from two voxels across sessions. Right inset, voxel activity correlated with the predicted activation from the logistic model after cross-validation. Positive r-values indicate increases in activity across sessions.

(B) Six-region parcellation of the lateral PFC in an example participant’s inflated left hemisphere. The lPFC was divided along a rostral-caudal and dorsal-ventral axis by combining smaller parcels from a multi-modal atlas of the cerebral cortex (Glasser et al., 2016).

(C) Left: mean activity for each fMRI session during the WM delay period for reliably active voxels (within each session), thresholded at t > 2.5. The dorsal rostral PFC ROI (green) showed a mean decrease in WM delay activity across sessions. Right: mean activity for all voxels (unthresholded). For visualization, all ROIs with significant logistic model fits after FDR correction are indicated with a bolded plot border, along with the fitted logistic curve across sessions. No other ROIs showed a mean change in WM delay activity over the course of training.

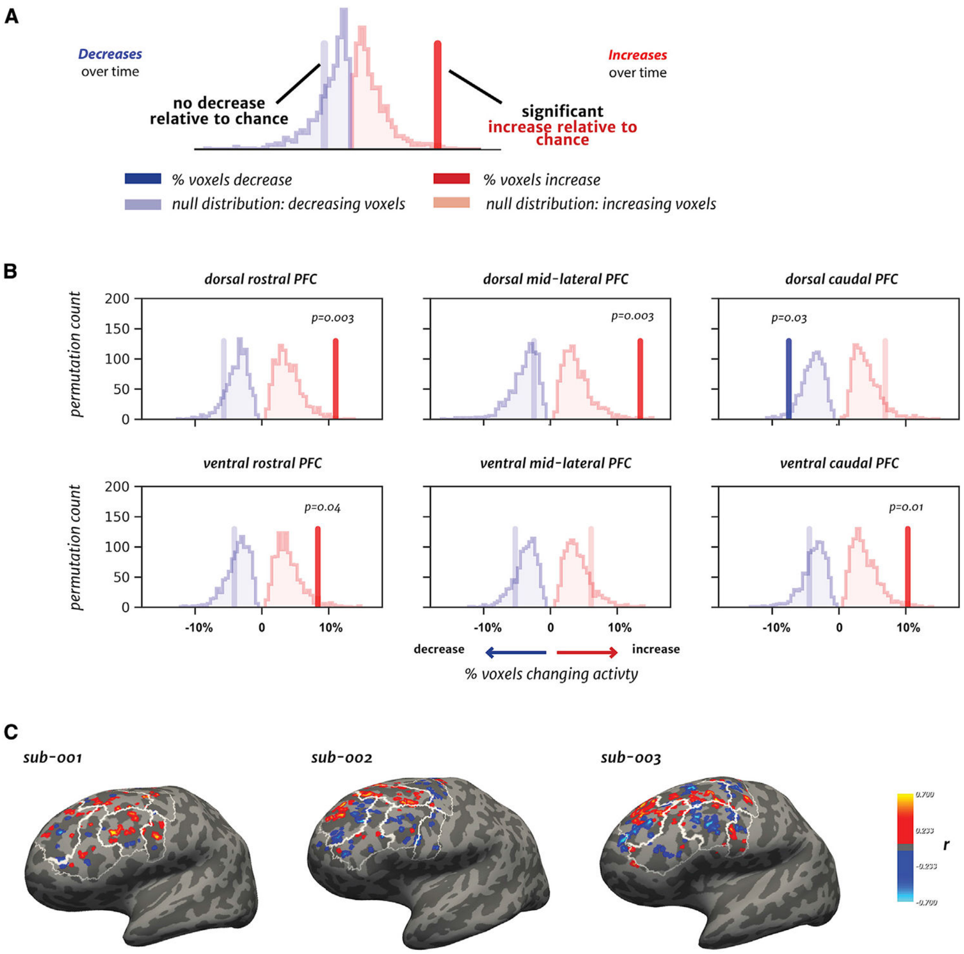

Figure 3. Distribution of WM delay activity patterns in the PFC across the course of learning

(A) Schematic of distribution of the percentage of voxels with increases (red; r > 0) or decreases (blue; r < 0) in activity across training (schematic). Significant changes over time are indicated by bolded vertical lines. Null distributions (created by permuting session number in the voxel-wise models) are shown in light red and blue.

(B) The percentage of voxels with significant changes in activity levels across training within each of the six lPFC ROIs. Four ROIs show a significant proportion of voxels with an increase in activity, while only the dorsal caudal PFC also shows a significant proportion of activity decreases.

(C) Voxel-wise maps of activity changes (thresholded at $p < 0.05$) for each participant plotted on their left hemisphere, inflated cortical surface, with the r value representing the correlation between the predicted and actual activation values shown in Figure 2A. Greater increases are shown by warmer colors, while decreases are shown with cooler colors.

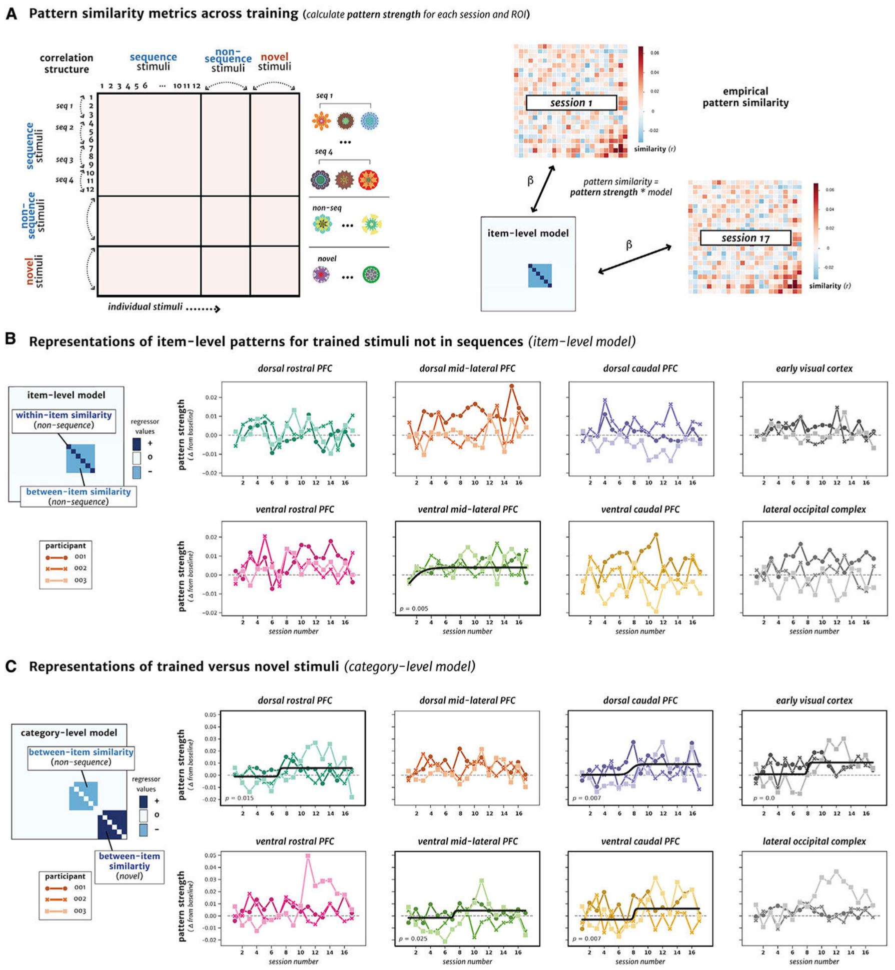

Figure 4. Changes in representational similarity patterns for trained items in WM delay activity

(A) Left: schematic of a WM delay activity pattern similarity matrix across different stimuli. Right: calculation of the pattern strength metric for each ROI and session by regressing a pattern model against the empirical pattern similarity data.

(B) Left: schematic of pattern similarity framework for the item-level model, where an interaction between on- (dark blue, positive values) versus off-diagonal (light blue, negative values) correlations among non-sequence trained stimuli serves as a measure of item-level representation. Right: plots of the pattern strength across sessions for each ROI, as assessed by the model fit for the on-versus off-diagonal interaction. For visualization, all ROIs with significant changes in pattern strength across sessions (after FDR correction) are indicated with a p value and bolded plot border. Pattern strength is plotted and fit with logistic models after subtracting baseline values (STAR Methods). Each line shading color and dot style represents one of the three individual participants.

(C) Same as in (B), but instead testing the category-level model for an interaction between trained (light blue, negative values) versus novel (dark blue, positive values) off-diagonal stimulus correlations.

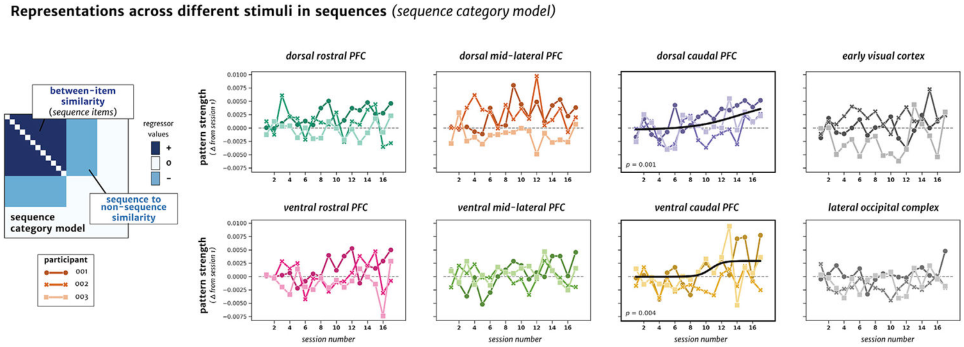

Figure 5. Changes in a categorical sequence representation in WM delay activity

Left: schematic of the model matrix for pattern similarity between items within trained sequences (dark blue, positive values) compared with trained items not in sequences (light blue, negative values). Right: plots of the pattern strength for each ROI, at each session, as assessed by the model fit for the sequence category model on the left. For visualization, all ROIs with significant changes in pattern strength across sessions (after FDR correction) are indicated with a p value and bolded plot border. Pattern strength is plotted and fit with logistic models after subtracting baseline values (STAR Methods). Each line represents one of the three individual participants.

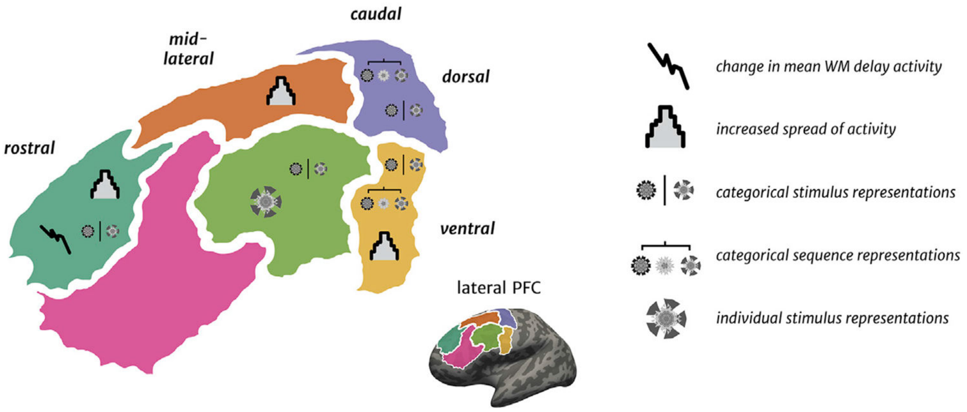
caudal
mid-lateral
dorsal
rostral
ventral
lateral PFC
change in mean WM delay activity
increased spread of activity
categorical stimulus representations
categorical sequence representations
individual stimulus representations

Figure 6. Summary of results

Left: each lPFC region, with icons depicting which WM delay activity metrics showed training-related changes. Right: legend for the symbols depicting significant changes in mean WM delay activity magnitude, activity spread within regions, or multivariate representations (from pattern similarity analyses) for sequences and items in WM.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and processed MRI data	This paper	OpenNeuro: ds003659; https://openneuro.org/datasets/ds003659
Analyzed activity maps	This paper	NeuroVault: https://neurovault.org/collections/12687/
Software and algorithms		
Python	Anaconda	https://www.anaconda.com/products/distribution
Freesurfer	Fischl et al. (1999a, b)	https://surfer.nmr.mgh.harvard.edu/
fMRIprep	Esteban et al. (2019)	https://fmriprep.org/en/stable/
Nilearn	Abraham et al. (2014)	https://nilearn.github.io/stable/index.html
Custom analysis code	This paper	Open Science Framework: https://doi.org/10.17605/OSF.IO/CKYBW https://osf.io/ckybw/?view_only=bd3c5ec6511a48c2b0a4f3329777cb16