Visual statistical learning is associated with changes in cortical manifold structure

Keanna Rowchan[1*], Daniel J. Gale[1], Qasem Nick[1,2], Jason P. Gallivan[1,2,3,‡], Jeffrey D. Wammes[1,2,‡]

[1]Department of Psychology, [2]Centre for Neuroscience Studies, and [3]Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada, K7L 3N6.

* Corresponding author
‡ The last two authors contributed equally to this work.

ABSTRACT

Our ability to implicitly extract environmental regularities over time and use this information to make predictions about the world is central to human cognition and behavior. While prior work has established the role of individual brain regions in this statistical learning (SL) process, it remains poorly understood how these regions coordinate their activity with the rest of the brain to guide SL. Here, using functional MRI, we measured human brain activity during visual SL, whereby individuals implicitly learned to associate pairs of images that were covertly embedded in a larger sequence of images. By projecting patterns of cortical and subcortical functional connectivity onto a low-dimensional manifold space, we found that SL was characterized by the expansion of several cortical areas along the whole-brain manifold. This included regions in higher-order visual cortex and the dorsal attention (DAN) and default-mode networks (DMN), and reflected their increased segregation from other brain networks. Consistent with these neural changes being associated with SL, we found that once the paired image structure was interrupted, these same brain areas immediately contracted back towards their pre-learning positions along the manifold. Notably, during this phase we also found that several new areas, belonging to the DAN and DMN, exhibited significant manifold contraction, with a subset of regions actually contracting beyond their baseline, pre-learning manifold locations. This suggests a putative mechanism in higher-order cortex by which learned associations can be represented over longer time horizons. Together, these results offer a new perspective on the distributed changes in whole-brain activity during visual SL.

Keywords: statistical learning, manifold, brain connectivity, gradients, hippocampus

Significance Statement

Here we used functional MRI and manifold learning techniques to explore changes in whole-brain cortical organization during visual statistical learning (SL). We found that areas in higher-order visual cortex and regions in frontoparietal cortex belonging to the dorsal attention network (DAN) and default mode network (DMN) exhibited significant changes in connectivity as individuals established implicit associations between pairs of visual items. Consistent with this effect being learning-dependent in nature, we found that once the pair structure was interrupted, these patterns of connectivity reverted back towards their baseline, pre-learning states. Notably, we found that this interruption also changed the connectivity of several new regions in the DAN and DMN, suggesting a differential role for these areas in the encoding learned associations.

INTRODUCTION

Each day we are bombarded with a constant influx of sensory information, and yet, despite these inputs seeming initially random, we are able to learn statistical patterns that emerge within our environment to guide future behaviour. Our ability to implicitly extract these environmental regularities over time is referred to as statistical learning (SL), and it allows us to make predictions about the world and modify our actions accordingly (1-3). One particularly well studied form of such learning is visual SL, which focuses on our implicit ability to generate internal representations of, and extract probabilities for, visual stimuli that tend to co-occur in the environment (e.g., in a temporal sequence or spatial array; 4-5).

At a neural level, prior work has implicated several areas within the subcortex and cortex in guiding visual SL. Within subcortex, areas of the medial temporal lobe (MTL; 6-8) have been most commonly associated with SL, a result that is consistent with the hippocampus' established role in learning (9-11) and its sensitivity to temporal and spatial information (12-16). Within cortex, both areas within the visual system and higher-order association cortex (e.g., prefrontal cortex (PFC), inferior frontal gyrus (IFG); 17-19) have been associated with SL, consistent with this form of learning not only requiring stimulus-specific processing (e.g., visual; 20-21, or auditory regions; 22-25, depending on stimulus), but also a more task-general ability to shift attention and cognitive resources (26-28). Importantly, while these previous studies have identified many of the component areas that contribute in isolation to SL (17-18; 22), it is widely appreciated that such learning is likely to involve the coordination and integration of information processing across multiple brain regions (see 29-31). Indeed, while prior work points to the involvement of both the hippocampus and visual cortex in visual SL, these brain areas project widely across the cortex and are reciprocally interconnected with several distributed, whole-brain networks (32-33). Understanding these regional interactions requires the study of functional connectivity between brain areas in large-scale networks, as well as analytical approaches that are able to characterize changes in whole-brain functional structure that occur during learning.

Recent advances in dimension reduction techniques have provided an emerging method for effectively studying and characterizing large-scale patterns of functional coupling among many distributed brain areas (34-37). These methods allow high-dimensional patterns of functional connectivity to be characterized along a low-dimensional subspace or manifold, which reflects the main patterns of covariance across brain regions. This approach has not only provided fundamental new insights into the principles that govern the collective activity of large-scale neural populations in many brain areas (38-39), but has also recently shed light onto the intrinsic functional organization of the human brain in both healthy and patient populations (40-42). More recently, this approach has been extended to explore temporal changes in whole-brain cortical organization that occur as subjects perform various tasks (34; 43-44). Here, we adopted this manifold learning approach to understand how cortical and subcortical activity patterns are coordinated during various stages of visual SL. Specifically, we projected human

subjects' patterns of functional connectivity onto a low-dimensional, connectivity-derived manifold, and examined changes in the relative positions of brain regions along the manifold prior to and during learning, as well as after learning, when learned associations were interrupted. This approach allowed us to capture the evolving landscape of brain activity during SL and understand how brain regions collectively re-organize their interactions over time to adapt to changing patterns in our environment.

RESULTS

The current study analyzed pre-existing functional magnetic resonance imaging (fMRI) data (45) collected from human participants (N = 33) performing a classic visual SL task, in which they viewed a sequence of abstract images (see Fig. 1A) presented one at a time on screen. The task was divided into eight blocks, where participants viewed a sequence of 80 images consisting of 5 repetitions of 16 unique images, resulting in a total of 640 image presentations. During six of these task blocks (blocks 2-7), the image sequences were structured to contain an embedded set of regularities, where the 16 images were combined into 8 pairs, such that one image in a pair (A) was always followed by its pairmate (B; See Fig. 1A, 1B left). These SL blocks were bookended by two, unstructured task blocks (blocks 1 and 8), where images were presented in a completely random, un-structured order (see Fig. 1B right). Importantly, participants were not explicitly aware of any embedded sequence structure throughout the task, and were only instructed to engage in a basic perceptual task that required the maintenance of attention, which was completely orthogonal to the pair structure.

Statistical learning results in neural changes at the regional level

Before examining and interpreting any changes in functional connectivity across cortical and subcortical regions during SL, we first sought to verify that subjects learned temporal associations between stimuli on the SL blocks. Prior work, including our own, has shown that visual SL leads to changes in stimulus-specific patterns of neural activity in the hippocampus (6-8, 45-46). Specifically, depending on the visual similarity between the paired images, learning can lead to either neural integration of representations, whereby the two stimuli that are temporally coupled become more similar in their neural representations, or to neural differentiation, whereby they become less similar. Recent theoretical (47-48), modeling (49-50) and empirical (45; 51-52) work indicates that this integration-differentiation relationship follows a U-shaped function (as described by the nonmonotonic plasticity hypothesis (NMPH: 46; 49; 53)). In our prior work based on these data (45), the strength of this U-shaped function in the dentate gyrus (DG) subfield of the hippocampus provided the primary measure that learning had taken place. Consistent with this prior finding, when we performed a re-analysis on the current data set including the subjects used for our functional connectivity analyses (see Methods), we observed a U-shaped learning effect in the CA1 ($r = 0.11$, 95% CI = [-0.010, 0.232], $p = 0.0351$) and DG ($r = 0.11$, 95% CI = [-0.017, 0.248], $p = 0.0451$), but not in the CA2/3 ($r = 0.07$, 95% CI = [-0.044, 0.180], $p = 0.115$), or across the

entire hippocampus when subfields were combined (r = 0.07, 95% CI = [-0.032, 0.180], p = 0.088; see Fig. 1C). These results replicate prior effects (45) and demonstrate that visual SL induced the hypothesized representational changes in brain structures known to support such learning. Building on this foundational result, we next examined whether, and to what extent, functional interactions both within and outside of the MTL system are also altered during SL.

Investigating changes in functional brain organization during statistical learning

To investigate changes in whole-brain functional organization during the SL task, we selected a functional run for each of four key epochs of the task. First, we defined Early-learning and Late-learning epochs based on the first and last structured blocks (80 image presentations each) of the SL task, respectively. Second, we defined Pre-learning and Post-learning epochs based on the first and last unstructured blocks (runs 1 and 8), in which the images were presented in a completely random order. For each participant, we extracted the mean blood oxygenation level-dependent (BOLD) timeseries data for every cortical region pre-defined by the Schaefer-1000 parcellation (53), as well as for each subcortical MTL region using individual-specific parcellations sourced from the Automated Segmentation of Hippocampal Subfields (ASHS; 54). We then computed covariance matrices (representing functional connectivity between each pairwise set of regions) for each of the four task epochs (Pre-learning, Early-learning, Late-learning, and Post-learning). [For a brief overview of the methodology, see Materials and Methods, Fig. 2A; and for the full methodology, see the Supplementary Extended Materials and Methods].

Prior work indicates that individual differences in functional connectivity account for the majority of the variance in patterns of brain activity, and thus these can easily obscure any task-related changes in functional connectivity (34, 55-56). We have recently developed methods (a Riemannian manifold centering approach; 56) to mitigate the effects of these individual differences, thereby allowing task-related changes in brain structure to be more readily detected (see Methods and Materials; 57). To illustrate the impact of this centering method, we projected participants' covariance matrices both before and after this centering procedure using uniform manifold approximation (UMAP; 58). UMAP effectively reduces the high-dimensional covariance matrix data onto a 2D plane, thus providing a visualization of the similarities/differences between the covariance matrices. Consistent with prior work (34; 55-56), before centering we found that the covariance matrices mainly clustered according to subject-level identity (see Fig. 2B, left). However, after our centering procedure, this subject-level clustering was completely abolished, thereby allowing the possibility of teasing apart learning-related differences in functional brain structure (see Fig. 2B, right).

Next, to investigate changes in cortical and subcortical patterns of functional connectivity during SL, we used the centered matrices to estimate separate cortical-subcortical connectivity manifolds for each participant's Pre-learning, Early-learning, Late-learning, and Post-learning covariance matrices. Building

from prior work (36, 59-60), we converted each centered connectivity matrix into an affinity matrix before applying PCA to the entire sample's affinity matrices. This reveals a set of principal components (PCs), which offer a low-dimensional representation of the cortical and subcortical structure (i.e., a cortical-subcortical manifold). Finally, we aligned each individual manifold to a common Pre-learning template manifold, constructed using the mean of all participants' Pre-learning matrices (see Fig. 2A). This alignment to the Pre-learning manifold served two important purposes: First, it provided a common target for manifold alignment, such that all participants and epochs could be directly compared within a common neural manifold space. Second, it allowed us to directly examine learning-specific deviations away from this Baseline task architecture, thus enhancing our sensitivity to detect learning-related changes in manifold structure (34).

Cortical-Subcortical Manifold Structure during Pre-Learning

The top three principal components (PCs) of our Pre-learning manifold effectively describe the main axes of cortical and subcortical functional organization during the task (see Fig. 3A), as well as explain the majority of the variance (60.57%; see Fig. 3B). PC1 distinguishes the visual network and a subset of DAN areas (positive loadings in red) from other cortical regions (negative loadings in blue). PC2 separates somatomotor regions of the cortex (in red) from transmodal, DMN regions (in blue). Lastly, PC3 primarily distinguishes between medial areas of the visual cortex (in red) and areas within the DAN (in blue). Together, PC1 and PC2 differentiate visual and somatomotor regions from transmodal regions of the DMN (see Fig. 3C), approximating the brain's intrinsic functional architecture, as observed in prior work using resting-state data (36-37). This tripartite functional architecture is thought to reflect a crucial feature of cortical brain organization, with the shift from unimodal to transmodal regions of the brain signifying a global processing hierarchy that ascends from lower-order sensory and motor systems, to higher-order association areas of the cortex, such as the DMN (61). By contrast, PC3 appears to be task-specific in nature, as it distinguishes association regions of the DAN from more sensory-specific regions of visual cortex, previously implicated in visual SL (62; see Fig. 3C).

In order to characterize the relative positions of both cortical and subcortical regions within the Pre-learning manifold, and thus provide a basis for examining *changes* from these positions during learning and post-learning, we computed, for each brain region, a measure called 'manifold eccentricity' (34; 40; 63-64). We derived the eccentricity of each region by taking its Euclidean distance from the manifold centroid (coordinates 0,0,0; see Fig. 3D). Quantitatively, this eccentricity metric provides a multivariate index of each region's embedding in the 3D manifold space (see Fig. 3E). Intuitively, regions with greater eccentricity (farther from the manifold's centroid) can be interpreted as having more functional segregation from other brain networks, whereas regions with lower eccentricity (closer to the manifold's centroid) can be interpreted as having more functional integration with other brain networks. In direct support of these interpretations, we show that regional eccentricity is strongly correlated with

classic graph theoretic measures of functional segregation and integration (see Fig. S1; Supplementary Extended Materials and Methods).

Changes in cortical-subcortical manifold structure during statistical learning

Having established the Pre-learning cortical-subcortical manifold, as well as a measure of a region's displacement along this manifold (eccentricity), we next sought to determine which regions exhibited significant changes in manifold embedding across our four task epochs (Pre-learning, Early-learning, Late-learning, and Post-learning). To this end, for each region, we performed repeated measures ANOVAs on eccentricity across the four task epochs, and corrected for multiple comparisons using false-discovery rate correction (FDR; $q < 0.05$; 65). We found that 40 regions exhibited a significant main effect of task epoch that survived FDR correction (see Fig. 4A). These regions almost entirely belonged to three major networks: the visual network (16 regions), the DAN (17 regions) and the DMN (5 regions). There was also one region each that arose in the somatomotor and limbic networks, respectively (see Fig. 4B).

To characterize the above main effects for each region, we performed follow-up FDR-corrected ($q < 0.05$) paired-samples t-tests across the different epochs. To test whether any brain regions changed their eccentricity at the onset of the learning blocks (when participants were first exposed to structured runs after the first baseline unstructured run), we performed a Early-learning > Pre-learning contrast. Notably, no brain regions exhibited significant differences in eccentricity between these two task epochs. This null result serves as an important control, given that, at the time of the early Learning run, participants would not yet have had enough exposure to the pair structure to establish temporal associations.

Next, to examine whether brain regions changed their eccentricity from the onset of learning to late learning, we performed a Late-learning > Early-learning contrast. This contrast revealed a general pattern of eccentricity increase across the majority of the areas that had been identified through the repeated measures ANOVAs (see Fig. 4A). Note that we also observed a general increase in eccentricity across functional brain networks (when all regions, significant or not, were averaged within their respective brain networks, 66; see Fig. 4C), though these effects were most prominent in the visual network, DAN and DMN. Together, these results show that, as learning unfolded, these visual, DAN and DMN regions exhibited significant expansion along the cortical manifold and thus, became more segregated from other brain networks.

Although these neural changes from Early- to Late-learning are presumed to reflect the learning process, it is possible that these effects are merely attributable to the passage of time. That is, rather than being learning-related in nature, these neural changes may simply reflect a general habituation effect and/or decrease in focus or attention as the task progresses. Behavior from the cover task indicates that this is

unlikely to be the case, as participants actually displayed faster reaction times and higher accuracy later in the task, improving from both Early- to Late-Learning, and Pre- to Post-learning, *p*s < 0.05. But, to test whether these alternative explanations can account for our results, we examined whether the observed pattern of manifold expansion also extended into the final, Post-learning epoch (when the learned sequence of structured images was interrupted). Notably, when we directly contrasted the Post-learning > Late-learning epochs, we observed a marked reversal in the pattern of effects we had found from Early- to Late-learning. Specifically, during Post-learning we found that all of the same regions that had exhibited expansion along the manifold during learning, now exhibited contraction along the manifold (see Fig. 4D, top center). Notably, the one network of regions that appeared to contradict this general pattern of effects was the MTL, which instead exhibited a slight increase in eccentricity (see Fig. 4D, top center).

Although the implication of the reversal of effects observed from Late- to Post-learning is that each brain region contracted back towards its original, Pre-learning position on the manifold, we noticed that this reversal effect appeared to be stronger in some functional networks (e.g., DAN, see Fig. 4E polar plot). Further investigation of this effect revealed that several brain regions within the DAN and DMN which did not exhibit significant expansion along the manifold during learning, now significantly contracted during Post learning (see Fig. 4E, bottom). To highlight this effect, we directly compared the absolute magnitude change in eccentricity from Early-to-Late learning versus Late-to-Post learning for each of our 40 significant brain regions (see Fig. 4F). When we sorted these regions according to their magnitude of eccentricity change, we found that, generally, areas were biased to exhibit larger contractions during Post-learning (below the horizontal zero line in Fig. 4F), as opposed to exhibiting larger expansions during learning (above the horizontal zero line in Fig. 4F), with several of these regions only showing significant changes during the Post-learning epoch (represented by the rightmost areas in Fig. 4F). This highlights that, once the structured sequence of images was interrupted (during Post-learning), not only did the same brain regions that exhibited increases in functional segregation now re-integrate (i.e., exhibit contraction along the manifold), but also that *additional* association regions in the DAN and DMN, not previously engaged during learning, increased their integration with other brain networks.

The greater extent of eccentricity changes during Post-learning in several of the above regions make it possible that several of these areas may have actually contracted beyond their original, Pre-learning positions on the manifold. To examine this, we performed a direct contrast of Post-learning > Pre-learning (which both contained a random sequence structure), and found that a subset of areas within the DMN and DAN, including bilateral areas of the frontal eye fields (FEF) and superior parietal cortex, and select higher-order visual regions, indeed changed their positions along the manifold as compared to Pre-learning (see Fig. 4E, see also Fig. 4F for significant Pre vs. Post pairwise comparisons denoted at the bottom of the plot). These findings indicate that, for a few higher-order visual, DAN and DMN regions, their embedding in the whole-brain manifold was significantly altered following learning.

Changes in cortical-subcortical connectivity that underlie the manifold changes

While our eccentricity measure offers a multivariate index of a region's overall connectivity profile, it does not itself reveal the *specific* changes in underlying connectivity that are driving the eccentricity changes. To unpack these effects, we performed seed connectivity analyses using representative seed regions taken from both the left and right hemispheres of the visual network, DAN and DMN, which together capture the main patterns of significant effects observed with respect to manifold expansion and contraction from Early- to Late-learning, and from Late- to Post-learning, respectively (see Fig. 5, and for corresponding analysis of the left hemisphere, see Fig. S2).

For each region, we contrasted seed connectivity maps both between Late-learning and Early-learning epochs (Late-learning > Early-learning) and between Post-learning and Late-learning (Post-learning > Late-learning) epochs by computing region-wise paired-samples t-tests. This allowed us to characterize the learning effect, and the post-learning reset, respectively, producing contrast maps depicting the connectivity changes associated with each representative seed region. Note that in Figure 5, we display the unthresholded voxel-wise contrast maps (see Fig. 5A) to allow for a more complete visualization of a given region's connectivity changes. In addition, the corresponding spider plots depict the network-level changes in connectivity for each seed region (see Fig. 5B), and the corresponding line plots depict each region's corresponding changes in eccentricity across epochs (see Fig. 5C). Together, these plots provide a comprehensive description of the collective changes in whole-brain connectivity that contribute to the changes in regional eccentricity. (Note that, for the spider plots, we used the 17-network mapping in order to capitalize on the improved spatial precision compared to the 7-network mapping; 53, 65).

Our seed connectivity analyses revealed that the pattern of learning effects seen from Early-learning to Late-learning (i.e., expansions of the visual, DAN and DMN regions along the cortical-subcortical manifold) for each seed region was generally driven by decreased covariance with hippocampal subfields and with areas outside of their respective networks, as well as increased covariance with areas located within their networks (see Fig. 5A, left). For instance, for the visual seed region (Fig. 5A, top row), there were widespread decreases in connectivity with areas like somatomotor and transmodal cortex, whereas the most prominent increase in connectivity was with early visual cortex. Similarly, for the DAN seed region in the frontal cortex (Fig. 5A, middle row), there were widespread decreases in connectivity with transmodal cortex, whereas the most prominent increase in connectivity was with superior parietal regions of the DAN. By contrast, when comparing Late-Learning to Post-learning, the observed increases in intra-network connectivity strongly reversed, with regions now exhibiting increases in covariance with other regions outside of their respective networks, including the MTL (see Fig. 5A, right). Indeed, note that for each seed region in Fig. 5A, the whole-brain connectivity maps in the left column (Late-learning > Early-learning) and right column (Post-learning > Late-learning) are near mirror images of one another. When investigating these seed region connectivity changes at the level of functional networks (Fig. 5B),

we also observed strong evidence for this robust increase and then decrease in connectivity from Early- to Late-learning and from Late- to Post-learning, respectively (see Fig. 5B).

As highlighted in our previous section above (see Fig. 4E and F), it is also clear that some of our representative seed regions—particularly within the DAN and DMN regions—appear to 'overshoot' their original, Pre-learning eccentricity positions along the manifold during the Post-learning epoch (Fig. 5C). Together, these seed-based analyses allow us to characterize, at the level of functional connectivity, selective changes in movement of areas along the whole-brain manifold during learning, as well as their reversal during Post-learning. (Note that although our seed connectivity analyses include only one region from each of the visual networks, DAN and DMN, we found that these findings were qualitatively similar regardless of which region was selected from each network).

The role of the hippocampus in statistical learning

In the previous sections, we have characterized regional changes in manifold eccentricity across higher-order visual, DAN and DMN areas during learning, as well as some of the underlying changes in functional connectivity that may be driving these manifold effects. Noticeably absent from these results, however, were statistically significant changes in the eccentricity of MTL regions across our task epochs (Pre-learning, Early-learning, Late-learning and Post-learning). One possible explanation for this discrepancy is that our FC analyses contained 998 cortical regions, and just 14 MTL regions. Since dimension reduction approaches like PCA are designed to maximize the amount of variance explained in a relatively small number of orthogonal components, it is not surprising that these components were dominated by changes in the cortex, and not hippocampus. Nevertheless, prior work has identified a clear role for substructures of the hippocampus in visual SL (6-8), and there are promising qualitative trends in our data to suggest the involvement of these areas during learning (e.g., consistent increases in eccentricity in the MTL across all of our pairwise comparisons, see Fig. 4C-E, top). Thus, here, we sought to more directly examine learning-related changes in the connectivity of hippocampal regions with cortex, as well as with each other. To this end, we derived seed-based hippocampal connectivity maps for subfields in left and right CA1, CA2/3, dentate gyrus (DG) and subiculum for each task epoch, and then created contrast maps to compare Late-learning to Early-learning, Post-learning to Late-learning and Post-learning to Pre-learning epochs.

Our hippocampal-based seed connectivity analysis revealed that from Early- to Late-learning, there was a strong decrease in covariance with cortical regions both within and outside the MTL, which was consistent across the whole-brain, regardless of network (see Fig. 6A, left). By contrast, when comparing Late- to Post-learning, we found that this trend reversed, such that the hippocampus now increased its covariance with other MTL areas and the cortex more broadly. This reversal in connectivity during Post-learning suggests increased functional communication between the hippocampus and cortical networks when the

learned pattern of structured sequences has been disrupted (see Fig. 6A, middle), possibly representing the propagation of a prediction error signal. Finally, when comparing the Pre- to Post-learning epochs, we observed that after learning, there was increased covariance across the whole-brain and MTL, relative to before learning had begun (see Fig. 6A, right).

To further specify the hippocampal-cortical interactions, we tested for learning-related changes in functional connectivity between bilateral hippocampus, and the representative cortical seed regions used in our prior analyses (see Fig. 6B for the three right cortical seed regions; Fig. S2 for the three left cortical seed regions). We averaged the correlations for each of these six seed regions across hippocampal regions (CA1, CA2/3, DG, and subiculum) within each hemisphere (6 cortical seed regions by 2 hippocampal hemispheres; 12 total pairings). These 12 possible pairings were each submitted to a one-way repeated measures ANOVA, looking for differences across task epochs, revealing a significant effect of task epoch across all ANOVAs (all p-values < 0.01). This was driven by significant decreases in hippocampal connectivity with our representative seed regions from Early- to Late-learning epochs (all p-values < 0.05). However, as soon as the learned associations were violated (during Post-learning), we observed the largest significant change in hippocampal-cortical connectivity between epochs (all p-values < 0.05), potentially reflecting the communication of hippocampal prediction errors to cortex. In addition, we also found a significant pattern of change in hippocampal-cortical connectivity from Pre- to Post-learning (all p-values < 0.05), suggesting that functional interactions between hippocampus and cortex were significantly altered following the learning process (for visualization of the left hippocampus, see Fig. S3).

Given prior work implicating the hippocampus in visual SL (67-68), it was surprising that our main eccentricity analyses did not identify significant changes in the positioning of hippocampal subregions along the whole-brain manifold across learning epochs. However, our targeted seed-based connectivity analyses offer insight into the whole-brain network organization with respect to the hippocampus's role in learning. Our results show a dramatic learning-related decrease in connectivity between the hippocampus and cortex, which is quickly reversed once learning was interrupted (in Post-learning), with the hippocampus exhibiting widespread increases in connectivity. These results provide a novel, whole-brain perspective to the hippocampal-cortical interactions that guide learning and later prediction.

DISCUSSION

The current study aimed to characterize changes in the landscape of whole-brain functional activity that underlie visual SL. We leveraged recent advances in dimension reduction techniques to build a connectivity-driven cortical-subcortical manifold space, which we used to characterize changes in regional functional organization across different stages of a visual SL task. We found that regions within higher-order visual cortex, the DAN and DMN exhibited expansion along the manifold during learning,

reflecting their functional segregation from other brain networks, and integration within their own networks. Consistent with the nature of these effects being learning-related, we found a reversal in this pattern of effects (i.e., contraction along the manifold) once the learned associations were interrupted, driven by the same visual, DAN and DMN areas re-integrating with other distributed networks. Notably, we also found that several additional regions that did not show strong learning effects — all within the DAN and DMN — exhibited significant contraction along the manifold, with several of these regions actually contracting beyond their baseline, Pre-learning positions. Finally, guided by previous work, we explored specific hippocampal-cortical interactions during the SL task and found that hippocampal subfields decreased their connectivity with cortex during learning, but then quickly re-integrated with cortex and MTL areas once the learned associations were interrupted. Together, these findings provide a unique window onto the whole-brain changes that underlie not only visual SL, but also the later violation of those statistically learned associations, and suggest key roles for higher-order visual and association regions in organizing learning over time.

From Early- to Late-learning, we found that higher-order regions within the visual network, DAN, and DMN exhibited significant expansion along the cortical-subcortical manifold. Our seed connectivity analysis indicated that, for representative regions from each network, this expansion was driven partially by decreased connectivity with other cortical networks and the MTL, including the hippocampus. This finding is consistent with prior work on SL showing reduced connectivity between visual (i.e., LOC) and DMN (i.e., precuneus) regions, as well as higher-order association regions in the frontal cortex (18). Similarly, while during initial learning hippocampal predictions are conveyed to visual cortex to entrain to presentation frequency (69-71) and sharpen their stimulus representations (72-76), once learning has been achieved, hippocampus (and other predictive regions) tend to reduce their activity and connectivity with sensory regions (18; 77-83). This suggests that as regularities are gradually encoded and predictions are developed and reinforced, cross-talk between distributed brain systems becomes less important for learning. Beyond these *decreases* in connectivity as learning proceeded, we discovered that the whole-brain manifold changes were also characterized by *increased* within-network functional connectivity. This suggests that the decrease in connectivity between areas observed in prior work is, at least in part, a consequence of increased within-network connectivity (i.e., modularity) within specific brain networks (e.g., higher-order visual cortex, the DAN and DMN). The changes we documented here may reflect a mechanism whereby once learning has taken place, the modularity of particular networks is increased in order to insulate the encoded associations from additional change (84-88), in turn enhancing the robustness of the learning.

A novel feature of our SL task was that the learning phases were book-ended by unstructured, random runs, allowing us to uniquely investigate changes in whole-brain activity patterns when the associations learned during visual SL were violated (i.e., from Late- to Post-learning). Consistent with the notion that the network segregation effects described above are related to visual SL, we found a striking reversal in

the manifold effects from Late- to Post-learning. Specifically, we found that every region that had exhibited significant expansions along the manifold during learning, now exhibited significant contraction along the manifold during Post-learning. Our seed connectivity analyses revealed that this contraction was driven by decreased within-network connectivity of the higher-order visual, DAN and DMN areas, and by their increased connectivity (integration) with the rest of the brain. This stark reversal in connectivity, driven in part by an increased covariance between these areas with the hippocampus, is likely related to a prediction error, or mismatch signal. Prediction error signals are typically evoked in the hippocampus (89-91), propagating out to cortical regions implicated in perception and attention (92-93) to update existing schemas (94). When predictions are violated, the outcome is often renewed communication between sensory and association areas (95), including between hippocampus and its sensory inputs (96-97). Our findings extend this theoretical view of SL, while providing a comprehensive, whole-brain perspective of how widely such prediction errors may be distributed across the cortex.

One particularly noteworthy finding in our study was that several regions that did not initially show learning effects—located solely within the DAN and DMN—now exhibited significant contraction when the learned associations were violated during the last, unstructured run (i.e., Post-learning). Notably, several of these regions, as well as a few of the higher-order visual areas, actually exhibited contraction *beyond* their initial, pre-learning positions on the manifold (see Fig. 4F). The majority of these areas were housed in association cortex, potentially indicating a more durable, long-lasting representation of the encoded memory. Indeed, other studies on implicit learning in several different domains (motor learning: 98-99, spatial-statistical learning: 100-103) have demonstrated that participants maintain a representation of learned statistical associations and regularities (e.g. motor commands, navigation locations) over long time periods, and will readily retrieve these representations to support behaviour days, and even months later (104). While higher-order cortical areas have been implicated in the longer-term storage of such memories and predictions (e.g.. consolidated and schematic learning; 90; 105-113), our current findings suggest a putative neural mechanism of how such memories might be encoded—namely, via shifts in embedding of individual regions within a whole-brain manifold space.

Changes in the positioning of MTL areas along the manifold were not obvious in our eccentricity analyses. Nevertheless, because prior work has implicated hippocampus and its substructures in visual SL (6-8; 45; 79) and later prediction (69; 81; 114-115), and because there were subtle hints that MTL behaved differently from the other networks (see Fig. 4D), we conducted seed connectivity analyses targeted at the hippocampus and MTL. We found that, during Early- to Late-learning, the hippocampus decreased its covariance not only internally, with other subfields, but also externally with other cortical networks. Then, when the learned associations were violated, this pattern completely reversed, with the hippocampus broadly increasing its connectivity with nearly every region of the brain. These results complement our characterization of the cortical consequences of prediction error, while also highlighting the dynamic

interactions between the hippocampus and the rest of the brain that support new learning and updating in response to change.

MATERIALS AND METHODS

We re-analyzed data from a previously published study (45) of 33 healthy individuals (20 female) between the ages of 18-35 years old, completing a visual SL task. Eight of the original dataset's 41 participants were excluded either because they completed fewer task runs than the majority of the sample (n = 7), or because there were issues with the quality of hippocampal and cortical segmentations (n = 1).

Participants completed eight experimental runs, where in each, 16 abstract images were presented one at a time on screen, each repeated five times, for a total of 80 image presentations. In all runs, the participants' unrelated cover task was to identify rare (10% of trials) instances of a small, faded grey square being overlaid on the images, by pressing a button. In the first experimental run (the Pre-learning epoch), the images were presented in a completely random, unstructured order. Following this run, participants completed six consecutive experimental runs (80 image presentations each, 480 total), in which unbeknownst to them, the 16 images were covertly combined into 8 pairs. This pair structure was embedded into the image sequence, such that a given image (A) was always followed by its pair-mate (B). The first and last of these structured runs were used as Early-learning and Late-learning epochs respectively, for the present analysis. Finally, after these structured runs, participants performed an additional Post-learning run, where they again viewed the 80 images in a completely random order (similar to the first run).

For our neural analyses, for each participant we obtained region-wise timeseries data by taking the mean blood oxygenation-level (BOLD) activity for each of the 998 cortical regions pre-defined by the Schaefer-1000 parcellation (two regions were removed due to small parcel size; 53) and each of 14 MTL regions, derived in an individual-specific way using ASHS software (54). Across the superset of these 1012 regions, we created centered functional connectivity matrices for each epoch, transforming them into cosine similarity affinity matrices after applying row-wise proportional thresholding. Next, we used PCA on affinity matrices to construct connectivity manifolds (35-36; 59-60) for each participant across epochs, aligning the principal components to a group-average template Pre-learning manifold using Procrustes transformation. To evaluate learning-related changes in manifold structure over time, we computed each brain region's manifold eccentricity (distance from 0,0,0) and assessed region-wise differences in eccentricity across task epochs using repeated measures ANOVAs and follow-up paired t-tests. To explore the alterations in whole-brain connectivity underlying observed changes in manifold eccentricity, we performed seed connectivity contrasts for three representative seed regions (belonging to the visual network, DAN and DMN) that demonstrated representative effects. To examine learning-related changes relative to hippocampus, we also performed similar seed connectivity contrasts for hippocampal

regions to examine changes during learning and post learning. For a comprehensive overview of the methods, please refer to the Supplementary Extended Materials and Methods.

Code availability. Imaging data were preprocessed using fmriPrep, which is open source and freely available. Regional timeseries extraction was performed using Nixtract, which is freely available in a repository at https://github.com/danjgale/nixtract. The analysis code is publicly available at https://github.com/thelamplab/SLmanifolds.

References

1. Saffran, J. R. (2009). What is statistical learning, and what statistical learning is not. Neuroconstructivism, 180-95.
2. Turk-Browne, N. B. (2012). Statistical learning and its consequences. The influence of attention, learning, and motivation on visual search, 117-146.
3. Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. Wiley Interdisciplinary Reviews: Cognitive Science, 8(1-2), e1373
4. Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. Journal of Experimental Psychology: General, 134(4), 552.
5. Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(2), 399.
6. Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. Journal of cognitive neuroscience, 26(8), 1736-1747.
7. Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. Current biology, 22(17), 1622-1627.
8. Schapiro, A. C., Turk‐Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. Hippocampus, 26(1), 3-8.
9. Mundy, M. E., Downing, P. E., Dwyer, D. M., Honey, R. C., & Graham, K. S. (2013). A critical role for the hippocampus and perirhinal cortex in perceptual learning of scenes and faces: complementary findings from amnesia and fMRI. Journal of Neuroscience, 33(25), 10490-10502.
10. Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. Journal of neurophysiology, 88(2), 982-990.
11. Preston, A. R., & Gabrieli, J. D. (2008). Dissociation between explicit memory and configural memory in the human medial temporal lobe. Cerebral Cortex, 18(9), 2192-2207.
12. Eichenbaum, H. (2017). Time (and space) in the hippocampus. Current opinion in behavioral sciences, 17, 65-70.
13. Geyer, T., Baumgartner, F., Müller, H. J., & Pollmann, S. (2012). Medial temporal lobe-dependent repetition suppression and enhancement due to implicit vs. explicit processing of individual repeated search displays. Frontiers in Human Neuroscience, 6, 272.
14. Howard, M. W., & Eichenbaum, H. (2015). Time and space in the hippocampus. Brain research, 1621, 345-354.
15. Turk-Browne, N. B. (2019). The hippocampus as a visual area organized by space and time: A spatiotemporal similarity hypothesis. Vision research, 165, 123-130.
16. Westerberg, C. E., Miller, B. B., Reber, P. J., Cohen, N. J., & Paller, K. A. (2011). Neural correlates of contextual cueing are modulated by explicit learning. Neuropsychologia, 49(12), 3439-3447.
17. Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. Nature neuroscience, 5(5), 485-490.
18. Karuza, E. A., Emberson, L. L., Roser, M. E., Cole, D., Aslin, R. N., & Fiser, J. (2017). Neural signatures of spatial statistical learning: Characterizing the extraction of structure from complex visual scenes. Journal of cognitive neuroscience, 29(12), 1963-1976.
19. Karlaftis, V. M., Giorgio, J., Vértes, P. E., Wang, R., Shen, Y., Tino, P., ... & Kourtzi, Z. (2019). Multimodal imaging of brain connectivity reveals predictors of individual decision strategy in statistical learning. Nature human behaviour, 3(3), 297-307.
20. Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. Journal of cognitive neuroscience, 21(10), 1934-1945.
21. Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. Journal of neuroscience, 30(33), 11177-11187.
22. Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. Brain and language, 127(1), 46-54.
23. McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. Journal of Neuroscience, 26(29), 7629-7639.

24. Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., ... & Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. Neuroimage, 48(3), 541-553.
25. Overath, T., Cusack, R., Kumar, S., Von Kriegstein, K., Warren, J. D., Grube, M., ... & Griffiths, T. D. (2007). An information theoretic characterisation of auditory encoding. PLoS biology, 5(11), e288.
26. Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities.
27. Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. Trends in cognitive sciences, 19(3), 117-125.
28. Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. Psychological Bulletin, 145(12), 1128.
29. Park, J., Janacsek, K., Nemeth, D., & Jeon, H. A. (2022). Reduced functional connectivity supports statistical learning of temporally distributed regularities. NeuroImage, 260, 119459.
30. Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews neuroscience, 10(3), 186-198.
31. Van Den Heuvel, M. P., Mandl, R. C., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. Human brain mapping, 30(10), 327-3141.
32. Barnett, A. J., Reilly, W., Dimsdale-Zucker, H., Mizrak, E., Reagh, Z., & Ranganath, C. (2020). Organization of cortico-hippocampal networks in the human brain. BioRxiv, 2020-06.
33. Froudarakis, E., Fahey, P. G., Reimer, J., Smirnakis, S. M., Tehovnik, E. J., & Tolias, A. S. (2019). The visual cortex in context. Annual review of vision science, 5, 317-339.
34. Gale, D. J., Areshenkoff, C. N., Standage, D. I., Nashed, J. Y., Markello, R. D., Flanagan, J. R.,... & Gallivan, J. P. (2022). Distinct patterns of cortical manifold expansion and contraction underlie human sensorimotor adaptation. Proceedings of the National Academy of Sciences, 119(52), e2209960119.
35. Paquola, C., Vos De Wael, R., Wagstyl, K., Bethlehem, R. A., Hong, S. J., Seidlitz, J., ... & Bernhardt, B. C. (2019). Microstructural and functional gradients are increasingly dissociated in transmodal cortices. PLoS biology, 17(5), e3000284.
36. Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., ... & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences, 113(44), 12574-12579
37. Huntenburg, J. M., Bazin, P. L., & Margulies, D. S. (2018). Large-scale gradients in human cortical organization. Trends in cognitive sciences, 22(1), 21-31.
38. Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. Annual review of neuroscience, 43, 249-275.
39. Shenoy, K. V., Sahani, M., & Churchland, M. M. (2013). Cortical control of arm movements: a dynamical systems perspective. Annual review of neuroscience, 36, 337-359.
40. Bethlehem, R. A., Paquola, C., Seidlitz, J., Ronan, L., Bernhardt, B., Tsvetanov, K. A., & Cam-CAN Consortium. (2020). Dispersion of functional gradients across the adult lifespan. Neuroimage, 222, 117299.
41. Hong, S. J., Vos de Wael, R., Bethlehem, R. A., Lariviere, S., Paquola, C., Valk, S. L., ... & Bernhardt, B. C. (2019). Atypical functional connectome hierarchy in autism. Nature communications, 10(1), 1022.
42. Xia, M., Liu, J., Mechelli, A., Sun, X., Ma, Q., Wang, X., ... & He, Y. (2022). Connectome gradient dysfunction in major depression and its association with gene expression profiles and treatment outcomes. Molecular Psychiatry, 27(3), 1384-1393.
43. Samara, A., Eilbott, J., Margulies, D. S., Xu, T., & Vanderwal, T. (2023). Cortical gradients during naturalistic processing are hierarchical and modality-specific. NeuroImage, 271, 120023.
44. Song, H., Shim, W. M., & Rosenberg, M. D. (2023). Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. Elife, 12, e85487
45. Wammes, J., Norman, K. A., & Turk-Browne, N. (2022). Increasing stimulus similarity drives nonmonotonic representational change in hippocampus. elife, 11, e68344.
46. Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. Nature neuroscience, 16(4), 486-492.
47. Frank, D., Montemurro, M. A., & Montaldi, D. (2020). Pattern separation underpins expectation-modulated memory. Journal of Neuroscience, 40(17), 3455-3464.
48. Ritvo, V. J., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic plasticity: how memory retrieval drives learning. Trends in cognitive sciences, 23(9), 726-742.

49. Ritvo, V. J., Nguyen, A., Turk-Browne, N. B., & Norman, K. A. (2023). Differentiation and Integration of Competing Memories: A Neural Network Model. bioRxiv, 2023-04.
50. Midler, B., & McClelland, J. L. (2023). Retrieval induced forgetting in a non-monotonic hippocampal model. bioRxiv, 2023-08.
51. Lewis-Peacock, J. A., & Norman, K. A. (2014). Competition between items in working memory leads to forgetting. Nature Communications, 5(1), 5768.
52. Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. Neuropsychologia, 51(12), 2371-2388.
53. Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., ... & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cerebral cortex, 28(9), 3095-3114.
54. Wisse, L. E., Kuijf, H. J., Honingh, A. M., Wang, H., Pluta, J. B., Das, S. R., ... & Geerlings, M.I. (2016). Automated hippocampal subfield segmentation at 7T MRI. American Journal of Neuroradiology, 37(6), 1050-1057.
55. Areshenkoff, C., Gale, D. J., Standage, D., Nashed, J. Y., Flanagan, J. R., & Gallivan, J. P. (2022). Neural excursions from manifold structure explain patterns of learning during human sensorimotor adaptation. Elife, 11, e74591.
56. Areshenkoff, C. N., Nashed, J. Y., Hutchison, R. M., Hutchison, M., Levy, R., Cook, D. J., ... & Gallivan, J. P. (2021). Muting, not fragmentation, of functional brain networks under general anesthesia. NeuroImage, 231, 117830.
57. Zhao, Q., Kwon, D., & Pohl, K. M. (2018). A riemannian framework for longitudinal analysis of resting-state functional connectivity. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11 (pp. 145-153). Springer International Publishing.
58. McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426
59. Vos de Wael, R., Benkarim, O., Paquola, C., Lariviere, S., Royer, J., Tavakol, S., ... & Bernhardt, B. C. (2020). BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. Communications biology, 3(1), 103.
60. Hong, S. J., Xu, T., Nikolaidis, A., Smallwood, J., Margulies, D. S., Bernhardt, B., ... & Milham, M. P. (2020). Toward a connectivity gradient-based framework for reproducible biomarker discovery. NeuroImage, 223, 117322.
61. Smallwood, J., Bernhardt, B. C., Leech, R., Bzdok, D., Jefferies, E., & Margulies, D. S. (2021). The default mode network in cognition: a topographical perspective. Nature reviews neuroscience, 22(8), 503-513.
62. Sherman, B., & Turk-Browne, N. (2022). Attention and Memory.
63. Park, B. Y., Bethlehem, R. A., Paquola, C., Larivière, S., Rodríguez-Cruces, R., Vos de Wael, R., ... & Bernhardt, B. C. (2021). An expanding manifold in transmodal regions characterizes adolescent reconfiguration of structural connectome organization. elife, 10, e64694.
64. Park, B. Y., Park, H., Morys, F., Kim, M., Byeon, K., Lee, H., ... & Bernhardt, B. C. (2021). Inter-individual body mass variations relate to fractionated functional brain hierarchies. Communications Biology, 4(1), 735.
65. Benjamini, Y., & Yekutieli, D. (2005). False discovery rate–adjusted multiple confidence intervals for selected parameters. Journal of the American Statistical Association, 100(469), 71-81.
66. Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology.
67. Tang, W., Christiansen, M. H., & Qi, Z. (2022). Human statistical learning dynamically shapes the hippocampal processing of temporal associations. BioRxiv, 2022-04.
68. Finn, A. S., Kharitonova, M., Holtby, N., & Sheridan, M. A. (2019). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. Journal of Cognitive Neuroscience, 31(1), 126-137.
69. Sherman, B. E., Aljishi, A., Graves, K. N., Quraishi, I. H., Sivaraju, A., Damisah, E. C., & Turk-Browne, N. B. (2023). Intracranial Entrainment Reveals Statistical Learning across Levels of Abstraction. Journal of Cognitive Neuroscience, 1-17.

70. Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., ... & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. Science Advances, 7(8), eabc4530.
71. Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. Cortex, 90, 31-45.
72. Ekman, M., Kusch, S., & de Lange, F. P. (2023). Successor-like representation guides the prediction of future events in human visual cortex and hippocampus. Elife, 12, e78904.
73. Tanaka, K. Z., Pevzner, A., Hamidi, A. B., Nakazawa, Y., Graham, J., & Wiltgen, B. J. (2014). Cortical representations are reinstated by the hippocampus during memory retrieval. Neuron, 84(2), 347-354.
74. Tompary, A., Duncan, K., & Davachi, L. (2016). High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. Hippocampus, 26(8), 995-1007.
75. Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. Proceedings of the National Academy of Sciences, 114(39), 10473-10478.
76. Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. Journal of cognitive neuroscience, 26(7), 1546-1554.
77. Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. Topics in cognitive science, 11(3), 482-503.
78. Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. Brain mapping, 3, 501-506.
79. Sherman, B. E., Graves, K. N., Huberdeau, D. M., Quraishi, I. H., Damisah, E. C., & Turk-Browne, N. B. (2022). Temporal dynamics of competition between statistical learning and episodic memory in intracranial recordings of human visual cortex. Journal of Neuroscience, 42(48), 9053-9068.
80. Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. Cerebral Cortex, 23(10), 2467-2478.
81. Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., ... & Maquet, P. (2008). Both the hippocampus and striatum are involved in consolidation of motor sequence memory. Neuron, 58(2), 261-272.
82. Rieckmann, A., Fischer, H., & Bäckman, L. (2010). Activation in striatum and medial temporal lobe during sequence learning in younger and older adults: relations to performance. Neuroimage, 50(3), 1303-1312.
83. Yon, D., Thomas, E. R., Gilbert, S. J., de Lange, F. P., Kok, P., & Press, C. (2023). Stubborn predictions in primary visual cortex. Journal of Cognitive Neuroscience, 35(7), 1133-1143.
84. Kirschner, M., & Gerhart, J. (1998). Evolvability. *Proceedings of the National Academy of Sciences*, *95*(15), 8420-8427.
85. Félix, M. A., & Wagner, A. (2008). Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity*, *100*(2), 132-140.
86. Kashtan, N., & Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences*, *102*(39), 13773-13778.
87. Wagner, G. P., & Altenberg, L. (1996). Perspective: complex adaptations and the evolution of evolvability. *Evolution*, *50*(3), 967-976.
88. Wimsatt, W. C., & Schank, J. C. (2004). Generative entrenchment, modularity and evolvability: When genic selection meets the whole organism. *Modularity in development and evolution*, 359-394.
89. Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. Journal of Neuroscience, 36(50), 12677-12687.
90. Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: the role of prediction and context. Trends in cognitive sciences, 19(2), 92-99.
91. Chen, J., Cook, P. A., & Wagner, A. D. (2015). Prediction strength modulates responses in human area CA1 to sequence violations. Journal of Neurophysiology, 114(2), 1227-1238.
92. Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. Progress in neurobiology, 192, 101821.
93. Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nature neuroscience, 19(5), 665-667.
94. Wang, S. H., & Morris, R. G. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. Annual review of psychology, 61, 49-79.
95. Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. Journal of Neuroscience, 29(43), 13445-13453.
96. Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron, 46(5), 703-713.

97. Bein, O., Plotkin, N. A., & Davachi, L. (2021). Mnemonic prediction errors promote detailed memories. Learning & Memory, 28(11), 422-434.
98. Hadjiosif, A. M., Morehead, J. R., & Smith, M. A. (2023). A double dissociation between savings and long-term memory in motor learning. *Plos Biology*, *21*(4), e3001799.
99. Krakauer, J. W. (2009). Motor learning and consolidation: the case of visuomotor rotation. *Progress in motor control: a multidisciplinary perspective*, 405-421.
100. Graves, K. N., Sherman, B. E., Huberdeau, D., Damisah, E., Quraishi, I. H., & Turk-Browne, N. B. (2022). Remembering the pattern: A longitudinal case study on statistical learning in spatial navigation and memory consolidation. *Neuropsychologia*, *174*, 108341.
101. Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit?. *Neuroscience letters*, *461*(2), 145-149.
102. Arciuli, J., & Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience letters*, *517*(2), 133-135.
103. Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322-1331.
104. Huberdeau, D. M., Krakauer, J. W., & Haith, A. M. (2019). Practice induces a qualitative change in the memory representation for visuomotor learning. *Journal of neurophysiology*, *122*(3), 1050-1059.
105. Tompary, A., & Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. Neuron, 96(1), 228-241.
106. Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. Trends in cognitive sciences, 21(8), 618-631.
107. Ruiter, D. J., van Kesteren, M. T., & Fernandez, G. (2012). How to achieve synergy between medical education and cognitive neuroscience? An exercise on prior knowledge in understanding. Advances in Health Sciences Education, 17, 225-240.
108. Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. Current opinion in behavioral sciences, 1, 1-8.
109. Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nature communications, 6(1), 8151.
110. Richards, B. A., Xia, F., Santoro, A., Husse, J., Woodin, M. A., Josselyn, S. A., & Frankland, P. W. (2014). Patterns across multiple memories are identified over time. Nature neuroscience, 17(7), 981-986.
111. Kuhl, B. A., & Chun, M. M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. Journal of Neuroscience, 34(23), 8051-8060.
112. Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., ... & Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. Nature neuroscience, 4(10), 1043-1048.
113. Hebscher, M., Wing, E., Ryan, J., & Gilboa, A. (2019). Rapid cortical plasticity supports long-term memory formation. *Trends in Cognitive Sciences*, *23*(12), 989-1002.
114. Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron, 60(2), 378-389.
115. Van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. Trends in neurosciences, 35(4), 211-219.

Figures

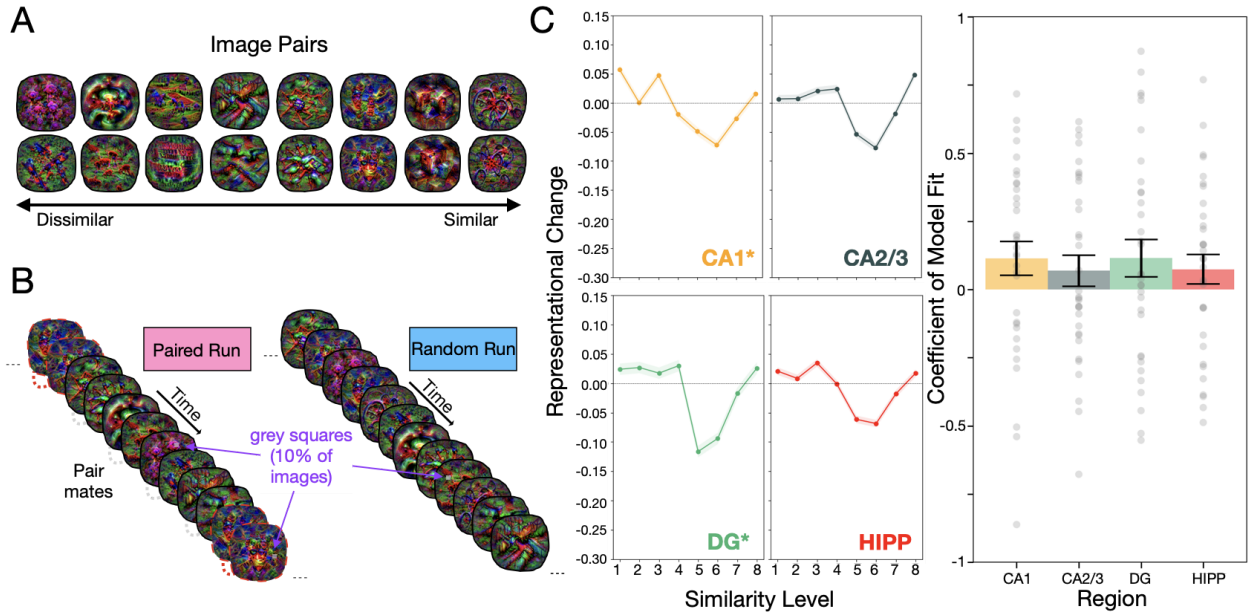

Figure 1. Overview of visual SL task and representational change predicted by the non-monotonic plasticity hypothesis. (A) Image pairs used in the visual SL task (45), ordered from most visually dissimilar, to most visually similar. (B) Visual SL task structure. Participants viewed images presented one at a time on screen. Structured runs contained an embedded pair structure, whereby an A image was always followed by its pairmate (B image; example of repeating pair mates denoted by red dashed line). By contrast, random runs presented images in a completely random order. Each run consisted of 80 image presentations (5 repetitions of each image or pair). Participants were tasked with pressing a button when a small faded gray square appeared in the image (10% of trials). (C) Left, Relationship between image similarity and representational change predicted by the nonmonotonic plasticity hypothesis (NMPH). For each region (CA1, CA2/3, DG and whole hippocampus; HIPP), we correlated parameter estimates derived from a general linear model (GLM) for each A and B image in a pair, and then created comparison scores by subtracting Pre-learning from Post-learning correlation scores. This measure quantifies the degree of representational change that takes place between image pairs as participants underwent visual SL (i.e., from Early- to Late-learning), with positive scores indicating integration between two images in a pair, and negative scores indicating differentiation. The comparison scores were plotted as a function of similarity level (i.e., coactivation) for each of the eight image pairs (ordered from least to greatest visual similarity in paired items). Shaded areas depict the bootstrap resampled standard errors of the means (SEMs). Right, theory-constrained cubic models were fit to the representational change patterns in all but one held-out participant, then used to predict the held-out participants' scores. These individual-specific model fits are shown in the light gray scatter plots. Barplots display the group-level model fit, as measured by the correlation between actual and predicted values. At the group-level, representational change was reliably predicted by a theory-constrained U-shaped model in the CA1 and DG. Error bars depict standard error of the mean (SEM).

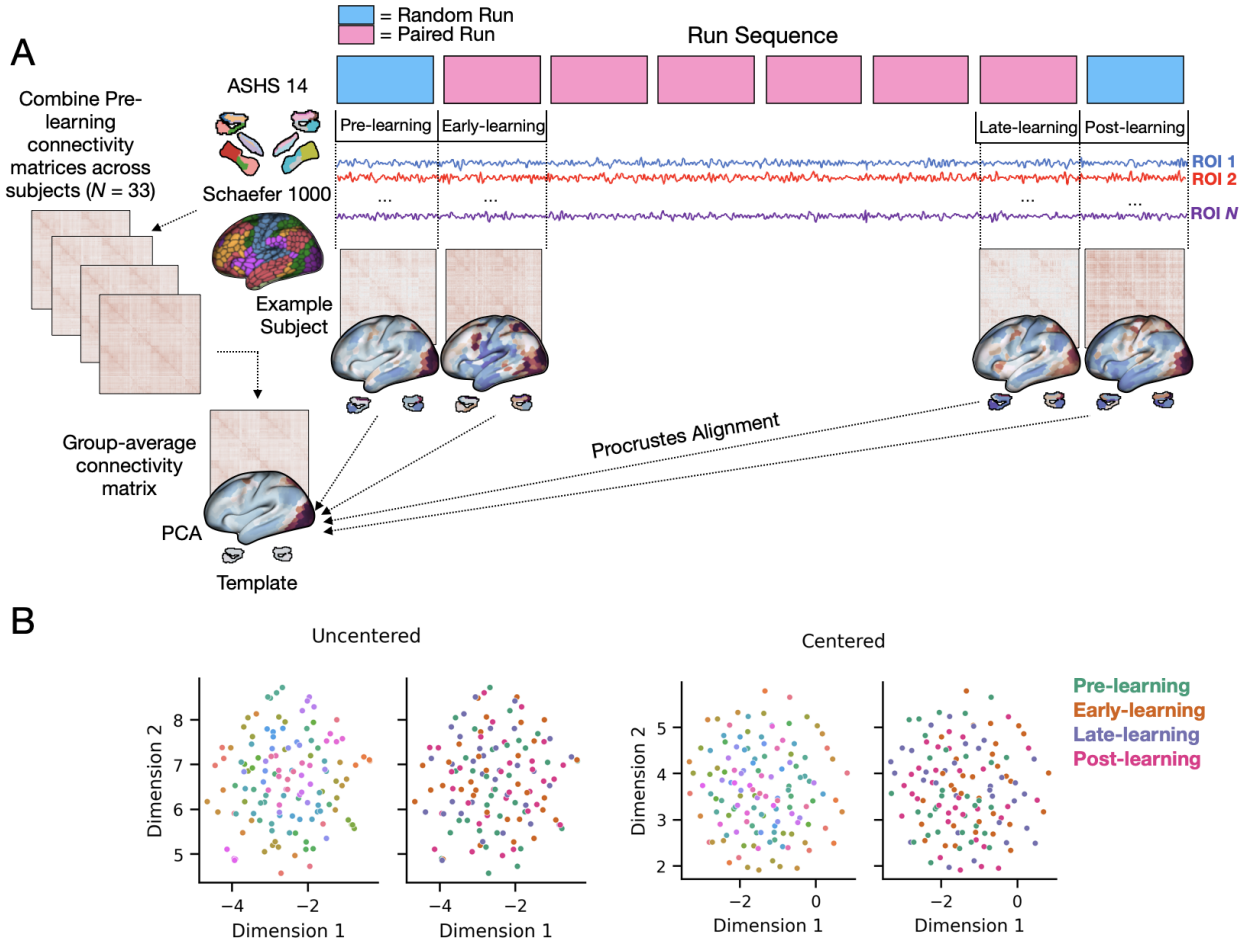

Figure 2. Overview of functional connectivity analysis. (A) Functional connectivity analysis approach. For each participant and task epoch (Pre-learning, Early-learning, Late-learning, and Post-learning), we computed functional connectivity matrices from parcel-wise time series data extracted from the Schaefer-1000 cortical parcellation (53) and using participant-specific hippocampal parcellations derived from the Automated Segmentation of Hippocampal Subfields (ASHS; 54) software. We then estimated functional connectivity manifolds for each task epoch using PCA using centered and thresholded connectivity matrices (see Methods). All manifolds (across all participants and epochs) were aligned to a common template manifold created from the group-average Pre-learning connectivity matrix using Procrustes alignment. (B) Uniform Manifold Approximation (UMAP; 58) visualization of the similarity of connectivity matrices, both before and after centering. Single data points represent individual connectivity matrices. For both the Uncentered and the Centered data, the left plots have the data points colored according to subject identity whereas the right plots have the data points colored according to task epoch. Note that prior to centering (Uncentered), the connectivity matrices were mainly clustered according to subject identity whereas this cluster structure was abolished following centering..

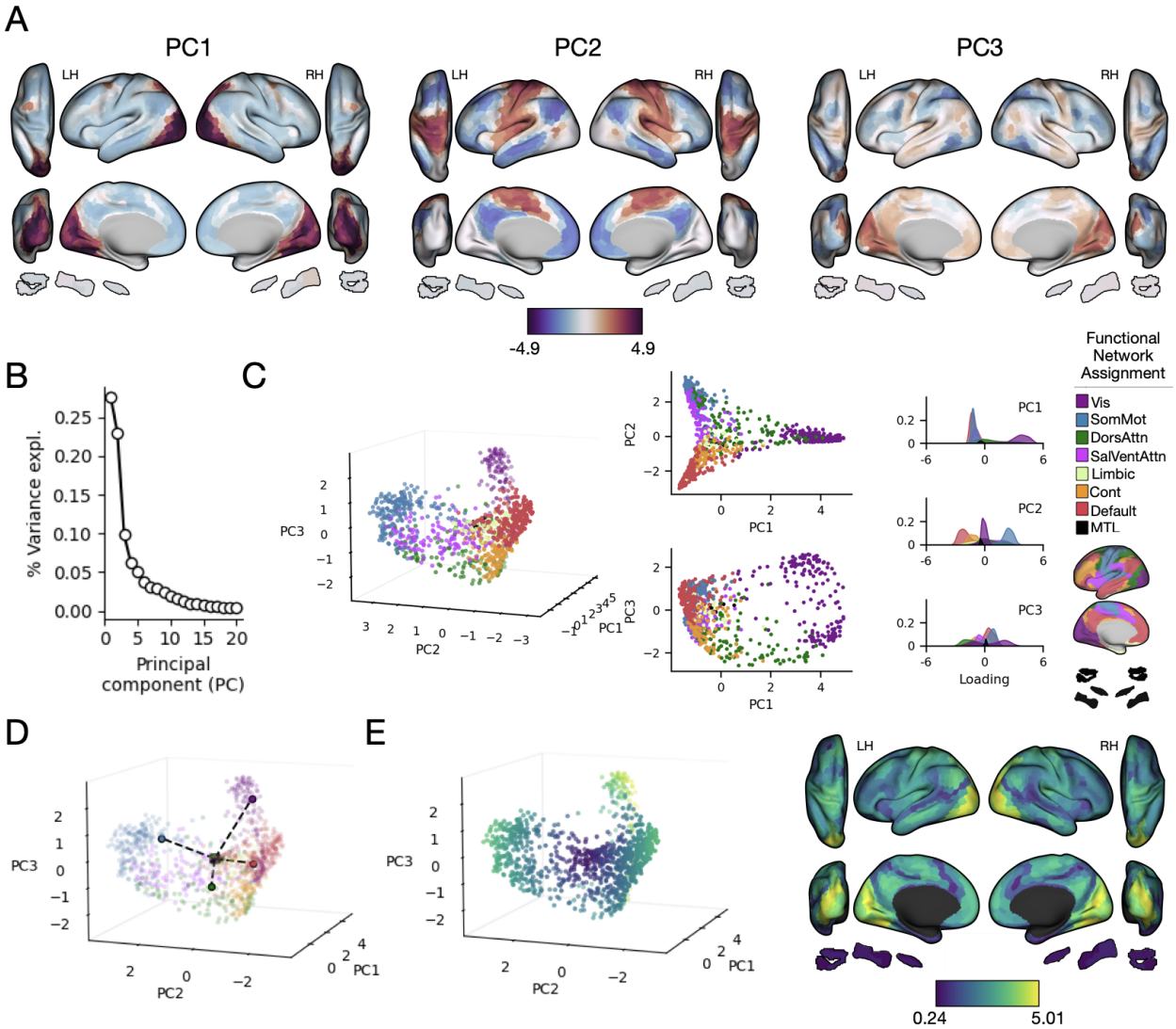

Figure 3. Manifold Structure and Eccentricity of Pre-learning Epoch. (A) Cortical-Subcortical loadings for each of the top three principal components (PCs), projected in surface space. (B) Percent variance explained for the first 20 PCs. (C) Functional organization of the Pre-learning epoch. Left, Each brain region is depicted as a point in a three-dimensional manifold space, with its location determined by its loading on the first three PCs, and its color determined by its functional network assignment, denoted at the right (66). Center, Each region is shown in two-dimensional spaces defined by PC1 and PC2 (top) and PC1 and PC3 (bottom). Right, histograms displaying the distributions of each network's loading along each PC. (D) Depiction of the Eccentricity calculation, derived by computing the distance between each region's position on the manifold, and the manifold centroid. (E) Eccentricity visualized in both manifold space (left) and surface space (right), where the color indicates the degree of eccentricity. Blue corresponds to low, and yellow to high, eccentricity.

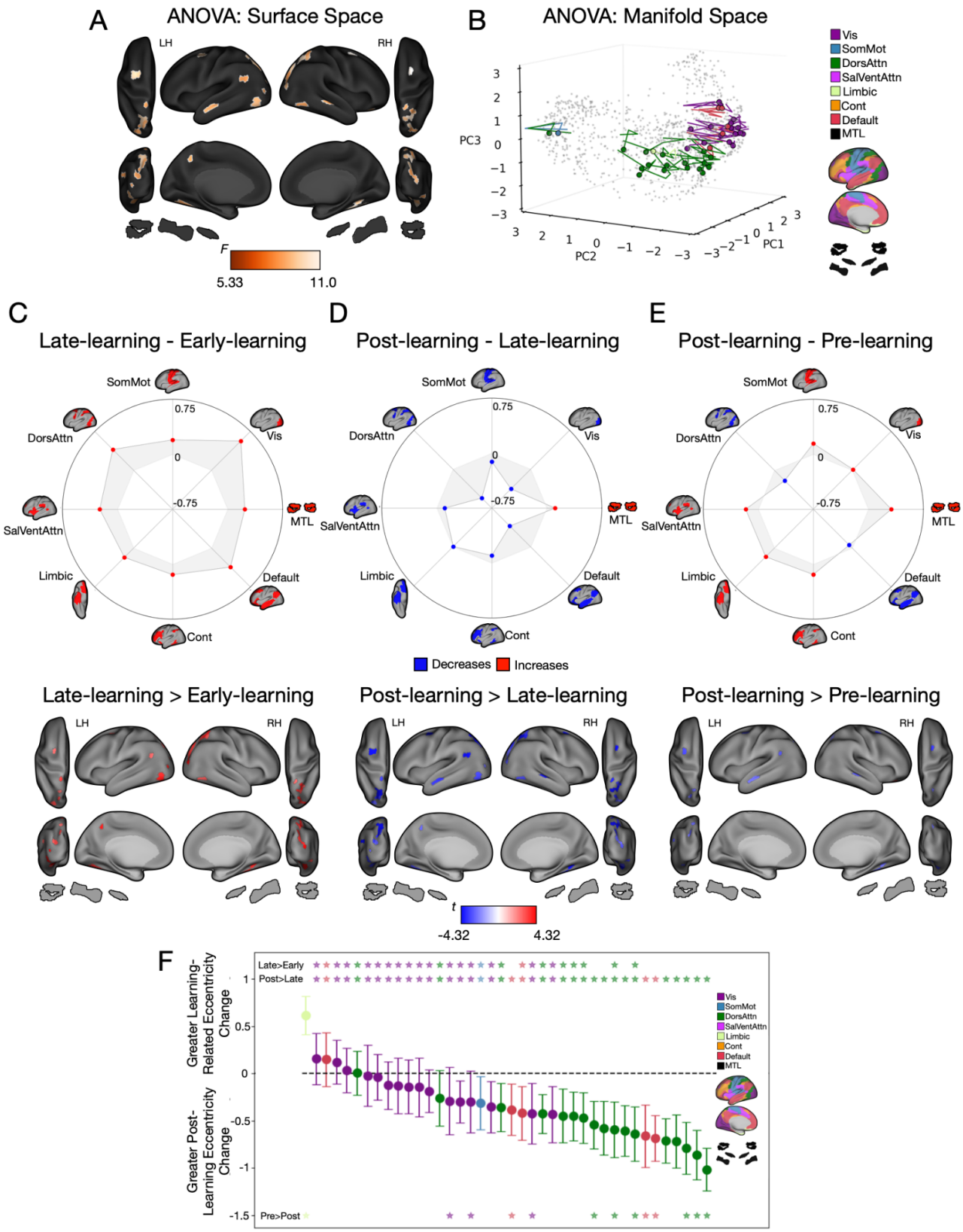
A
ANOVA: Surface Space
LH
RH
F
5.33
11.0
B
ANOVA: Manifold Space
PC3
PC2
PC1
Vis
SomMot
DorsAttn
SalVentAttn
Limbic
Cont
Default
MTL
C
Late-learning - Early-learning
D
Post-learning - Late-learning
E
Post-learning - Pre-learning
SomMot
Vis
MTL
Default
Cont
Limbic
SalVentAttn
DorsAttn
0.75
0
-0.75
Decreases
Increases
Late-learning > Early-learning
Post-learning > Late-learning
Post-learning > Pre-learning
t
-4.32
4.32
F
Greater Learning-Related Eccentricity Change
Greater Post-Learning Eccentricity Change
Late>Early
Post>Late
Pre>Post

Figure 4. Changes in manifold eccentricity across different phases of visual SL. (A) Regions displaying significant changes in eccentricity across epochs in our ANOVA analysis. (B) Temporal trajectories of the significant regions from A, projected in manifold space. Each region is depicted on the manifold as a point, in three-dimensional space, with significant regions coloured according to their respective functional network assignments on the right (66). Colored circles on significant regions denote the Pre-learning (Baseline) manifold epoch, and each trace attached to the circle represents that region's displacement on the manifold throughout the subsequent learning epochs (Early-learning, Late-learning and Post-learning). (C-E) Top, Mean functional network eccentricity differences for pairwise comparisons between epochs. Positive (red) dots reflect average increases along the manifold, and negative (blue) points show average contractions along the manifold. Bottom, Individual regions exhibiting significant pairwise comparisons between task epochs. Positive (red) values indicate increases in eccentricity and negative (blue) values indicate decreases in eccentricity. (F) Eccentricity differences for regions displaying significant changes in eccentricity in our ANOVA analysis, where regions have been sorted according to differences in their absolute magnitude of change during learning versus Post-learning. Scores below zero represent relatively greater contractions from Late- to Post-learning as compared to expansions from Early- to Late-learning and vice versa for scores above zero. Error bars represent the standard error of the mean (SEM). Horizontal stars along the top and bottom of the graph indicate which regions in our ANOVA were significant for each set of pairwise comparisons.

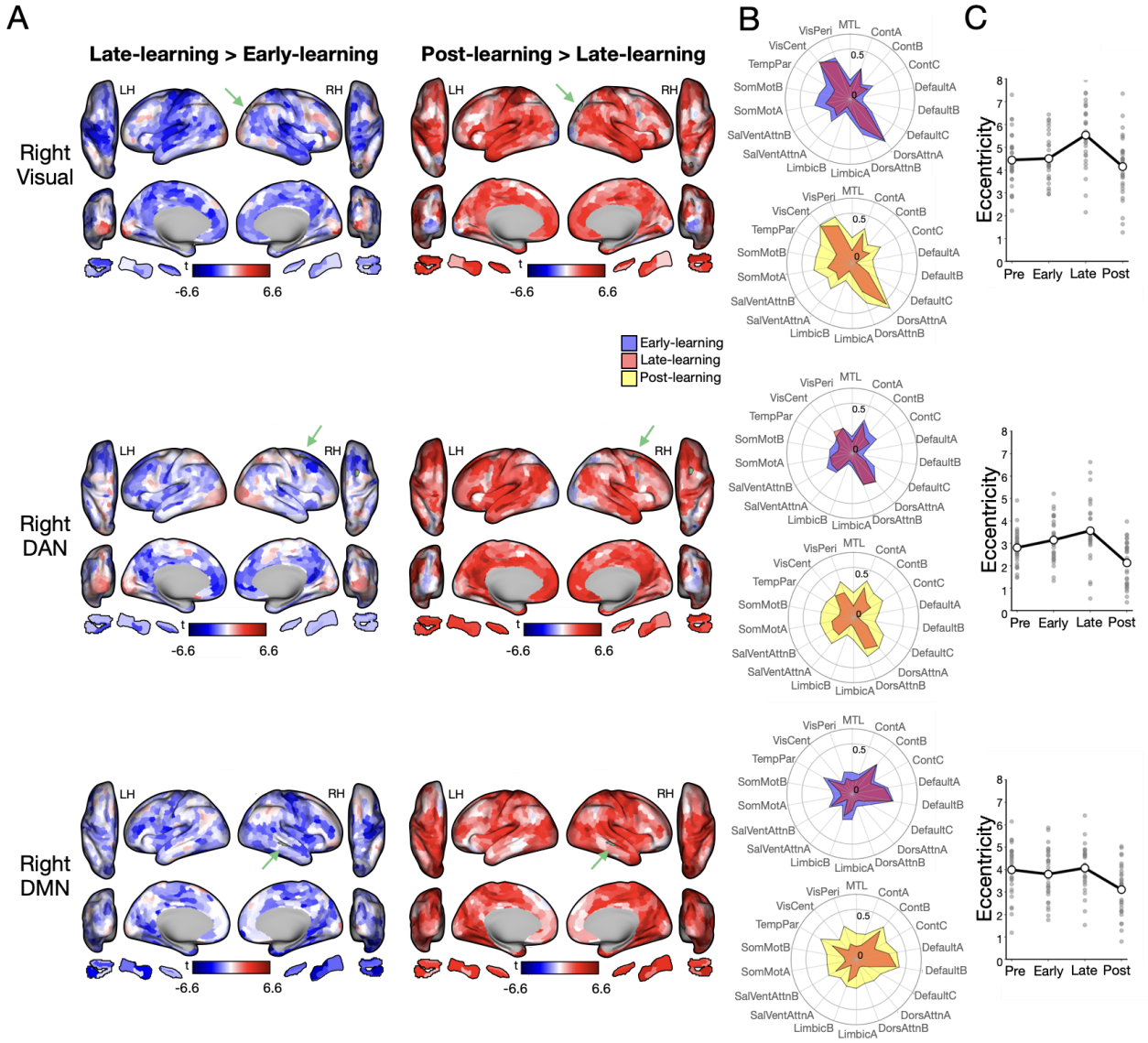

Figure 5. Patterns of connectivity underlying changes in eccentricity throughout visual SL. (A) Connectivity changes using representative seed regions in right visual, DAN and DMN areas (shown in green, as indicated by green arrow) from Late-learning>Early-learning and Post-learning>Late-learning. Positive (red) values show increases and negative (blue) values show decreases in connectivity. (B) Spider plots show seed-based changes in connectivity between epochs at the network level (according to the Yeo 17-networks parcellation (65) of the brain, as well as the additional MTL regions derived from the Automated Segmentation of Hippocampal Subfields (ASHS; 54) in black). (C) Changes in eccentricity for each seed region across task epochs (Pre-learning, Early-learning, Late-learning, and Post-learning). Large white dots denote the mean whereas gray dots denote single participants.

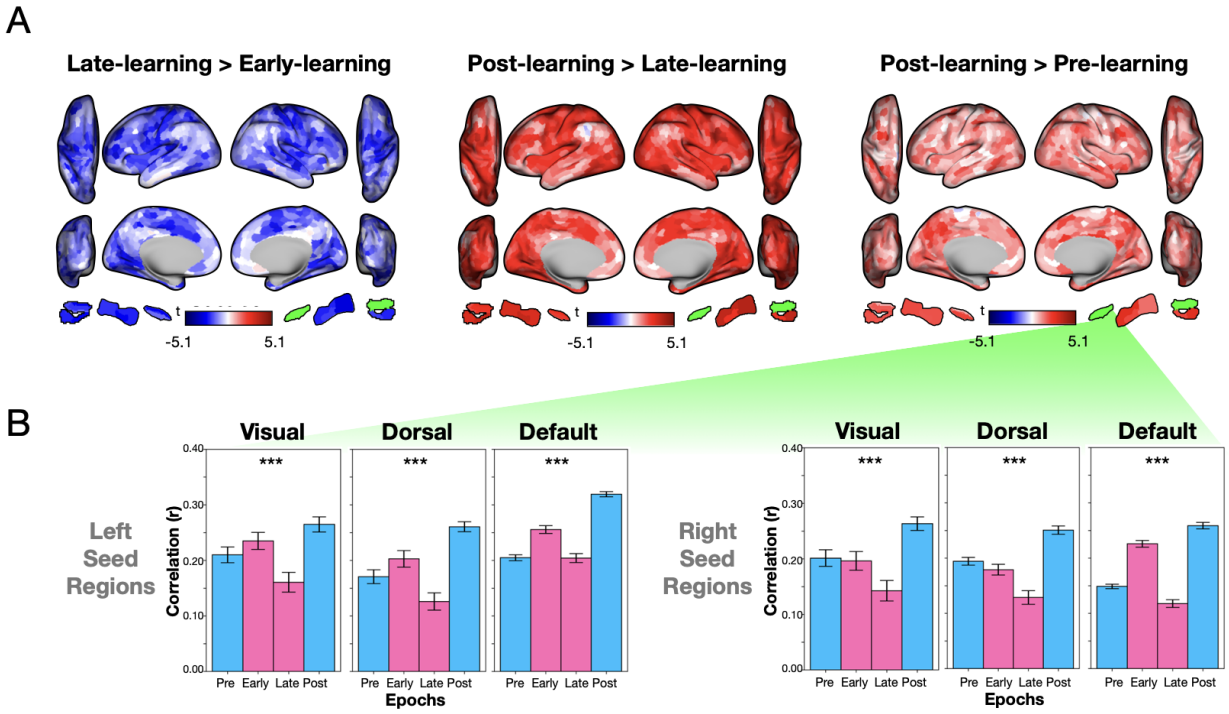

Figure 6. Hippocampal-Based Connectivity Analysis. (A) Connectivity changes for the right hippocampus (shown in green) from Late-learning>Early-learning, Post-learning>Late-learning, and Post-learning>Pre-learning. Positive (red) and negative (blue) values show increases and decreases in connectivity, respectively. (B) Correlation comparisons for aggregated right hippocampal regions with left and right seed regions in visual, DAN and DMN areas (from Fig. 5; Fig. S2). Error bars represent the within-hippocampal standard error of the mean (SEM), with significance asterisks indicating p-values for our ANOVA results $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Each significant ANOVA was followed up with paired samples t-tests. These t-tests revealed that the most significant changes in connectivity were between Early-learning to Late-learning, (all p-values < 0.05), Late-learning to Post-learning (all p-values < 0.05), and Pre-learning to Post-learning (all p-values < 0.05).