Dynamic flexibility in striatal-cortical circuits supports reinforcement learning

Raphael T. Gerraty[1], Juliet Y. Davidow[2], Karin Foerde[3], Adriana Galvan[4], Danielle S. Bassett[5,6], and Daphna Shohamy[1,7]

[1]Department of Psychology, Columbia University, New York, NY, USA.
[2]Department of Psychology, Harvard University, Cambridge MA, USA.
[3]Department of Psychology, New York University, New York, NY, USA.
[4]Department of Psychology, UCLA, Los Angeles, CA, USA
[5]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104 USA
[6]Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA
[7]Zuckerman Mind Brain Behavior Institute and Kavli Institute for Brain Science, Columbia University, New York, NY, USA.

Summary

Complex learned behaviors involve the integrated action of distributed brain circuits. While the contributions of individual regions to learning have been extensively investigated, understanding how distributed brain networks orchestrate their activity over the course of learning remains elusive. To address this gap, we used fMRI combined with tools from dynamic network neuroscience to obtain time-resolved descriptions of network coordination during reinforcement learning. We found that reinforcement learning involves dynamic changes in network coupling between the striatum and distributed brain networks. Moreover, we found that the degree of flexibility in whole-brain circuit dynamics correlates with participants' learning rate, as derived from reinforcement learning models. Finally, we found that episodic memory, measured in the same participants at the same time, was related to dynamic connectivity in distinct brain networks. These results support the idea that dynamic changes in network communication provide a mechanism for information integration during reinforcement learning.

Introduction

Learning from reinforcement is central to adaptive behavior and requires continuous and dynamic integration of sensory, motor, and cognitive information over time. Major progress has been made in understanding how individual brain regions support reinforcement learning. However, remarkably little is known about how these brain regions interact during learning, how their interaction changes over time, and how these dynamic circuit level changes relate to successful learning.

In a typical reinforcement learning task, participants use trial-by-trial reinforcement over hundreds of trials to learn to associate cues or actions with their most probable outcome (e.g. Daw, 2011; Frank et al., 2004; O'Doherty et al., 2003; Shohamy et al., 2004). Computationally, this sort of learning is captured by so-called "model-free" reinforcement learning algorithms, a class of models that provide a quantitative and mechanistic framework for describing behavior on a trial-by-trial basis (Daw, 2011; Daw et al., 2005; Sutton and Barto, 1998). Such models have also been highly successful in accounting for neuronal signals underlying learning behavior (Daw et al., 2006; O'Doherty et al., 2003; Schultz et al., 1997), collectively demonstrating a central role for the striatum and its dopaminergic inputs in model-free reinforcement learning. To accomplish such learning, however, the striatum and other regions must integrate visual, motor, and reinforcement information over time, a process that is likely to involve coordination across a number of different circuits.

The idea that the striatum serves such an integrative role in learning and cognition is not new (Bogacz and Gurney, 2007; Ding, 2015; Frank et al., 2015; Haber, 2003; Hikosaka et al., 2014; Kemp and Powell, 1971; Wiecki and Frank, 2013). Anatomically, the striatum is well positioned for such integration, given that it receives extensive input from many regions of cortex and projects back, through thalamus, to motor cortex (Alexander et al., 1986; Haber, 2003; Haber et al., 2006; Haber and Knutson, 2010). However, while the idea that the striatum serves such a role is anatomically and theoretically appealing, it has been very difficult to test this possibility empirically. Thus, whether the striatum interacts with other sensory, motor and cognitive regions during learning, and how such network-level interactions reconfigure over the course of learning, remains unknown.

Understanding the process of integration at the network level has been hampered by the lack of mathematical tools capable of elucidating those circuits in a data-driven fashion, and quantifying their engagement, integration, and flexible reconfiguration in real-time as humans adapt their behavior. Until now, most studies of large-scale brain connectivity have focused on static descriptions of networks (Bressler and Menon, 2010; Bullmore and Sporns, 2009; Damoiseaux et al., 2006; Power et al., 2011), limiting their ability to link networks to cognitive processes (Medaglia et al., 2015). But, there is increasing attention to the importance of a more dynamic perspective on circuit configuration. Even seemingly stable networks undergo temporal changes (Allen et al., 2012; Liu and Duyn, 2013;

Smith et al., 2012), implying that such static descriptions fail to capture transient patterns of co-activation that be essential for complex behavior.

Here we aimed to address this gap. We take advantage of recent advances in a dynamic formulation of graph theory and its application to neuroimaging data, an emerging field known as dynamic network neuroscience (Kopell et al., 2014; Medaglia et al., 2015). This formulation has been spurred by the development of algorithmic tools like multi-slice community detection (Mucha et al., 2010), which can be used to infer activated circuits and their reconfiguration from fMRI data collected as participants perform cognitively demanding tasks (Bassett et al., 2011; Bassett et al., 2015; Braun et al., 2015). Such tools have recently been leveraged to understand the role of dynamic brain-wide connectivity in motor skill learning (Bassett et al., 2011; Bassett et al., 2013b; Bassett et al., 2015). A key measure from this emerging field is an index of a brain region's tendency to communicate with different networks over time, known as "flexibility"(Bassett et al., 2011; Braun et al., 2015). Prior work has shown that flexibility across a number of brain regions predicts individual differences in the speed of acquisition of a simple motor task (Bassett et al., 2011). Network flexibility has also been shown to correlate with working memory and other dimensions of executive function (Braun et al., 2015). But their role in perhaps the most dynamic of cognitive behaviors – trial-by-trial updating of learning based on reinforcement – has not been studied.

Guided by the anatomical and computational considerations outlined above, we hypothesized that temporal network dynamics, and flexibility in particular, support key processes underlying reinforcement learning. Specifically, we hypothesized that (1) reinforcement learning involves flexible network coupling between the striatum and across brain circuits; (2) that these neural circuit changes would relate to measurable changes in behavior, specifically (i) *learning performance* (accuracy, within subjects) and (ii) *learning rate*, as derived from reinforcement models. We were particularly interested in the learning rate parameter because it quantifies the extent to which a learner uses reinforcement on any individual trial to update their responses. Thus, the learning rate is a good index of integration across trials: lower learning rates indicate a wider window of integration, while the extreme case of the highest possible learning rate (of 1) means that all updating happens based on a single trial, without any integration.

We also sought to explore the relationship between network flexibility and episodic memory for events that coincided with reinforcement, but which were incidental to the reinforcement learning task itself. Episodic memory is well known to depend on separate circuits in the medial temporal lobe. The rationale for testing episodic memory was two-fold. First, it provided an important control comparison for time-on-task effects on circuit dynamics. Second, it was a question of interest given that very little is known about how episodic memory is supported by dynamic circuit configuration in the context of ongoing reinforcement learning. Given the extensive literature indicating separate brain regions supporting episodic memory vs.

reinforcement learning, we hypothesized (3) that a distinct set of regions would exhibit a relationship between network flexibility and episodic memory.

To test these hypotheses, we used functional magnetic resonance imaging (fMRI) to measure changes in brain network structure while participants engaged in learning of probabilistic stimulus-outcome associations based on reinforcement (**Figure 1a**). On each trial, participants were presented with a visual cue, then made a choice indicated by a key press, and then received feedback. Associations between cues and outcomes were probabilistic and required updating based on trial-by-trial feedback. We used a task for which behavior has been well described by reinforcement learning models (Foerde and Shohamy, 2011), and which is known from fMRI to involve the striatum (Foerde and Shohamy, 2011) and from patient studies to depend on it (Foerde et al., 2012). To test the role of network dynamics in episodic memory, the task also included trial-unique images presented during feedback. Each of these images coincided with reinforcement, but they were incidental to the learning task (**Figure 1b**). Thus, this task allowed us to test hypotheses about how dynamic changes in network coupling relate to reinforcement learning and to episodic memory.

Results

Reinforcement Learning Performance. Participants learned the optimal response for each cue. Percentage of optimal responses increased continuously from 68% in the first block to 76% in the final block, on average. Using a mixed-effects logistic model predicting optimal response with learning block, we observed a significant effect of block on learning performance, as measured by the proportion of optimal responses during each block (**Figure 1c**; $\beta = 0.28$, Standard Error (S.E.) = 0.11, p = 0.01 (Wald approximation, (Bates et al., 2015)).

We also fit standard reinforcement learning models(Daw, 2011; Sutton and Barto, 1998) to participants' trial-by-trial choice behavior (**Experimental Procedures**). As expected, the models fit learning behavior significantly better than chance (**Experimental Procedures**, t_{21}=23.04, p<0.0001).

Of particular interest was the learning rate α, a parameter that indexes the extent to which an individual weighs feedback from single trials as opposed to averaging information across many trials. A low learning rate indicates that an individual is combining choice value over multiple experiences, and such a rate is preferable in an environment that is well characterized by static probabilities, such as the task used here (Sutton and Barto, 1998). The average learning rate was 0.50 with a standard deviation of 0.32. Consistent with prior studies and with the variability in learning accuracy (above), we also found that the learning rate varied across individuals, from 0.004 (high integration) to 0.966 (low integration), indicating that subjects varied considerably in the extent to which they integrated reinforcement across trials to update their responses. This variability allowed us to determine whether learning rate, across and within subjects, is related to differences in circuit dynamics during learning.

Brain Network Flexibility. To probe the role of dynamic brain circuits, we first sought to characterize spatial and temporal properties of brain network dynamics during the task, separately from any information about learning performance. We constructed dynamic functional connectivity networks for each subject in 50 s windows, and used a recently developed multi-slice community detection algorithm (Mucha et al., 2010) to partition each network into dynamic communities: groups of densely connected brain regions that evolve in time (**Experimental Procedures**). In 110 cortical and subcortical ROIs from the Harvard-Oxford atlas, we computed a flexibility statistic for each learning block, which measures the proportion of changes in that region's allegiance to large-scale communities over time (Bassett et al., 2011).

We began by examining the regional distribution of network flexibility across the brain during learning by computing the average flexibility across the whole learning task (see **Experimental Procedures**). As predicted, sensory and motor regions consistently showed the lowest levels of flexibility (Betzel et al., 2016), while association cortices showed moderate to high levels of flexibility (**Figure 2a**). In addition to this regional distribution, we sought to verify that flexibility was not related to the size of the ROI. We found that ROI size only explained 1.17 % of the variance in average flexibility across individuals, $r = 0.11$, $t_{108} = 1.311$, $p = 0.26$), indicating that this measure is not an artifact of the parcellation we used. We also studied the temporal profile of this measure, by examining changes in flexibility over the course of the task. Whole-brain flexibility (averaged across all ROIs) increased in early learning blocks, before decreasing in later stages of the task (**Figure 2b**).

Flexibility in the striatum relates to reinforcement learning. Previous reports have implicated large-scale network flexibility in motor performance. The temporal profile of network flexibility during this experiment suggests a parallel with learning behavior. To test specifically whether flexibility in the striatum's network coupling is related to learning performance, we fit a mixed effects logistic regression (Bates et al., 2015) using average flexibility in an *a priori* striatum ROI during individual learning blocks to predict performance. Striatal flexibility computed for each block was significantly associated with accuracy on the same block (**Figure 3a**; $\beta = 9.45$, S.E. = 2.75, $p < 0.001$ (Wald approximation (Bates et al., 2015))) indicating that network flexibility is related to behavioral measures of learning. For appropriate posterior inference, we fit a Bayesian extension of this model using Stan (Carpenter et al., 2015) (**Experimental Procedures**), to generate a posterior 95% credible interval (C.I.) of [3.53, 14.99], indicating that flexibility changes of 1% are associated with learning changes of 0.88-3.80% on average (**Supplemental Figure 1**). To ensure that this relationship reflected a within-subjects effect of flexibility on learning, we included average flexibility across blocks for each subject in the model. This extended model produced similar results ($\beta = 9.79$, S.E. = 2.82, $p = 0.0005$). These results indicate that changes in coupling between the striatum and other brain networks during learning play a significant role in reinforcement learning.

Flexibility relates to learning in a distributed set of brain regions. To explore regions other than the striatum whose dynamic connectivity across networks predicted learning, we performed the above analysis first for the whole brain and then for each of the 110 cortical and subcortical ROIs. These analyses also allowed us to explore the effects of dynamic network coupling in distinct sub-regions of the striatum, as we performed separate tests on the left and right caudate, putamen, and nucleus accumbens.

Consistent with reports relating whole-brain flexibility to behavior (Bassett et al., 2011; Braun et al., 2015), we found a significant effect of global flexibility averaged across all ROIs on learning (**Supplemental Figure 2**, $\beta = 11.84$, S.E. = 3.91, $p<0.005$). By testing each ROI separately, we found further evidence for our hypothesis that the striatum flexibly communicates with a number of brain networks in order to integrate stimulus-action-value information. This analysis again revealed a significant effect of flexibility in striatal sub-regions (the right putamen and left caudate) surviving False Discovery Rate (FDR) correction for multiple comparisons. While the global effect might have been driven solely by the striatum, other regions may also play a role. The corrected results, which are presented in **Figure 3b**, indicate that network flexibility in regions of the motor cortex, parietal lobe, and orbital frontal cortex, among others (see **Supplemental Table 1 and Figure 3** for full list and uncorrected map), are also associated with learning from probabilistic feedback.

Individual differences in learning rate correlate with whole-brain flexibility. We next explored the relationship between flexibility and the learning rate α, a parameter in computational models of reinforcement learning, that quantifies the extent to which subjects weigh feedback from each trial when computing the current value of an action given a particular stimulus (Daw, 2011; Sutton and Barto, 1998), as described above. We focused on flexibility averaged across the whole brain and in the *a priori* striatum ROI. Consistent with our hypothesis and with the observation of wide variability in learning rate across subjects, learning rate exhibited a significant negative correlation with whole-brain flexibility (Spearman's correlation coefficient $\rho = -0.60$, $p<0.005$, **Figure 4**); that is, participants with a lower learning rate (indicating more integration of information across multiple trials) had more flexibility. There was no significant relationship between learning rate and the proportion optimal responses ($\rho = -0.28$, $p = 0.21$), suggesting that this correlation reflected learning per se, and not general performance. Moreover, the relationship between learning rate and flexibility remained significant while controlling for task performance (partial $\rho = -0.62$, $p < 0.005$). We did not detect a relationship between flexibility in the striatal ROI and learning rate (**Supplemental Figure 4,** $\rho = 0.004$, $p>0.95$).

Together these results indicate that reinforcement learning involves dynamic changes in network structure, especially in the coupling of distributed regions that may each be involved in different aspects of learning. They also suggest that

individual differences in learning and updating can be explained in part by dynamic coupling across brain networks.

Flexibility in medial cortical regions is associated with episodic memory. The task also included trial-unique objects presented simultaneously with reinforcement, allowing us to measure subjects' episodic memory, a process thought to rely on distinct cognitive and neural mechanisms to feedback-based incremental learning. We tested whether network flexibility was associated with episodic memory for these trial-unique images, as assessed in a later surprise memory test. Having a measure of episodic memory for the same trials in the same participants allowed us to determine whether the same striatal network dynamics are correlated with any form of learning, or whether these two forms of learning, occurring at the same time, are related to distinct network dynamics.
Behaviorally, participants' memory was better than chance (d-prime = 0.93, t_{21} = 7.27, $p < 0.0001$). Memory performance ("hits") varied across learning blocks, allowing us to assess within-subject associations between network flexibility and behavior (**Figure 5a**). Memory performance was not correlated with incremental learning performance (mixed effects logistic regression, $\beta = 0.41$, Standard Error (S.E.) = 0.51, p = 0.42 (Wald approximation)). We tested the effect of flexibility on memory performance (proportion correct) in each of the 110 ROIs. A whole-brain FDR-corrected analysis revealed one region where flexibility was associated with episodic memory, the left paracingulate gyrus. An exploratory uncorrected analysis revealed regions in the medial prefrontal and medial temporal (parahippocampal) cortices where flexibility was associated with episodic memory ($p<0.05$ uncorrected, **Figure 5b**). None of the sub-regions from our *a priori* striatum ROI passed even this low threshold for an effect of flexibility on episodic memory.

Discussion

Learning through repeated reinforcement is essential for adaptive behavior. Learning relies on the dynamic reconfiguration of brain circuits to integrate across multiple domains, combining and maintaining information about sensory, motor, value, and memory to guide decisions. Until recently, understanding the changes in brain networks related to learning has been challenging. Most studies have focused on the characterization of individual regions or static patterns of connectivity (Gerraty et al., 2014; Harmelech et al., 2013; Tambini et al., 2010), largely due to a lack of available methods for interrogating the role of network dynamics as behavior evolves.

The current results provide evidence that reinforcement learning, long linked to a value updating signal originating in the midbrain (Schultz et al., 1997), involves dynamic coordination of learning-related regions such as the striatum with a set of broader cortical networks. Our central findings, that network flexibility is associated with learning performance as well as single-trial updating parameters from reinforcement learning models, are consistent with the hypothesis that this coordination underlies the brain's ability to integrate information over time to assign value to an action in the presence of a particular cue.

An interesting hypothesis generated by the current framework is that network flexibility will play a larger role in learning behavior the more that the learning in question relies on widespread information integration, and that this flexibility will be selective to the regions known to participate in this type of learning. Thus, instrumental conditioning paradigms involving complex audio-visual stimuli (Kehoe and Gormezano, 1980), for example, would be expected to associate more strongly with striatal flexibility than the task presented here. In addition, learning that relies on inferences or integration (Wimmer et al., 2012; Wimmer and Shohamy, 2012) or on the encoding of associations across space or time (Eichenbaum, 2000; Staresina and Davachi, 2009), both processes thought to rely on medial temporal and prefrontal processes, should be associated with network flexibility in these regions. This framework provides testable hypotheses to be explored in future work.

Multiple disorders have been linked to both learning abnormalities (Fineberg et al., 2010; Foerde et al., 2015; Kheirbek et al., 2012; Waltz and Gold, 2007) and changes in diffuse connectivity patterns (Greicius, 2008), with increasing evidence that these changes involve aberrations in connectivity dynamics (Damaraju et al., 2014; Ma et al., 2014). The results presented in this report raise the possibility of linking these separate strands of research to better understand mental diseases with learning components, such as schizophrenia (Waltz and Gold, 2007), anxiety (Kheirbek et al., 2012), and eating disorders (Foerde et al., 2015), among others. In this framework, abnormalities in the integration and routing of information, expressed as differences in dynamic patterns of large-scale functional connectivity (Braun et al., 2016), could lead to the deficiencies in learning hypothesized to represent cognitive endophenotypes for mental disorders (Robbins et al., 2012).

There are a number of limitations to this report. First, given the static feedback probabilities, the relationship between network flexibility and learning performance could be confounded by the amount time on task for each subject. This seems unlikely to fully explain this relationship, given that the relationship between episodic memory and network flexibility did not increase linearly over time. Nonetheless, future studies incorporating reversal periods to disassociate overall time on task from learning performance will further address this issue. Another limitation is the hard partition for network assignments provided by current multi-slice community detection methods utilized in this study, which necessarily underemphasize uncertainty about community assignment. There have been very recent attempts to formalize probabilistic models of dynamic community structure (Durante et al., 2016; Palla et al., 2016), but most work examining dynamic communities in the brain have used deterministic community assignment (Bassett et al., 2011; Bassett et al., 2015; Shine et al., 2016). More work needs to be done to develop and validate these probabilistic models and apply them to neuroscience data. Relatedly, while the spatial resolution of fMRI makes it an appealing method to begin examining network dynamics, further characterization will entail the use of methods with higher resolution such as ECoG (Khambhati et al., 2016) and MEG (Siebenhühner et al., 2013).

To summarize, we report a novel link between computational models of reinforcement learning, which have often relied on the activity of individual regions to provide an underlying brain mechanism for learning from repeated experiences, with a dynamic measure of network connectivity. Studies of the role of large-scale brain networks in behavior have in contrast often eschewed discussion of the mechanistic roles of single regions, in part because of the challenge of linking widespread connectivity patterns to specific temporal epochs (Medaglia et al., 2015). However, recent advances in network theory are beginning to make the role of dynamic communication between individual regions and broader networks in a host of behaviors a tractable area of research (Bassett et al., 2011; Bassett et al., 2015; Braun et al., 2015; Kopell et al., 2014). Incremental learning based on reinforcement, which necessarily relies on large-scale information integration to inform future behavior, represents a cognitive domain in which such dynamic local-widespread communication provides a rich framework for understanding brain-behavior relationships (Büchel et al., 1999).

Materials and Methods

Experimental Design. Twenty-five healthy right-handed adults (age 24-30 years, mean of 27.7, standard deviation of 2.0, 13 females) were recruited from University of California Los Angeles and the surrounding community as the adult comparison sample in a developmental study of learning (Davidow et al., 2016), and provided informed consent in writing to participate in accordance with the UCLA Institutional Review Board. Subjects were paid for their participation. Participants reported no history of psychiatric or neurological disorders, or contraindications for MRI scanning. Three subjects were excluded from this analysis (two for technical issues in behavioral data collection and one for an incidental neurological finding), leading to a final sample size of 22.

Task and Behavioral Analysis. The probabilistic learning task administered to subjects undergoing an fMRI session has been previously described (Davidow et al., 2016; Foerde et al., 2012; Foerde et al., 2013; Foerde and Shohamy, 2011). Before scanning, participants completed a practice round of 8 trials to become familiar with the task. During the imaging session, individuals underwent an instrumental conditioning procedure, in which they learned to associate 4 cues with two possible outcomes. The cues were images of butterflies; the outcomes were images of flowers. On each trial, participants were presented with an image of one of the four butterflies along with two flowers, and asked to indicate which flower the butterfly was likely to feed from, using a left or right button press. They were then given feedback consisting of the words 'Correct' or 'Incorrect'. Presentation of the feedback also included an image of an object unique to each trial, shown in random order for the purpose of subsequent memory testing. For each butterfly image, one flower represented the 'optimal' choice, with a 0.8 probability of being correct, while the alternative flower had a 0.2 probability of being followed by correct feedback. Subjects performed four blocks of this probabilistic learning phase, each consisting of 30 trials. Feedback was presented for 2 seconds, and was followed by a

randomly jittered inter-trial interval. These four blocks were followed by a test phase, in which subjects performed the same butterfly task without feedback for 32 trials.

For each trial in the learning phase, both the feedback received as well as whether or not subjects made the optimal choice were recorded, and percent correct for each block was computed as the percent of trials on which subjects made the optimal choice, regardless of feedback. These variables enable a characterization of learning as the proportion of optimal choices in each block, as well as that in the test phase. Using this information, we fit reinforcement learning models to subjects' decisions (Daw, 2011; Sutton and Barto, 1998).

Following the fMRI session (30 minutes), subjects were given a surprise memory test for the trial-unique object images presented during feedback in the learning phase. Subjects were presented with all 120 objects shown during the conditioning phase, along with an equal number of novel objects, and asked to judge the images as "old" or "new". They were also asked to rate their confidence for each decision on a scale of 1-4 (one being most confident; four indicating "guessing"). All responses rated 4 were excluded from our analyses (Foerde and Shohamy, 2011).

Reinforcement Learning Model. To characterize learning on a trial-by-trial basis, we fit standard reinforcement learning models to individuals' choice behavior (Daw, 2011; Sutton and Barto, 1998). Briefly, the expected value for a given choice at time t, Q_t, is updated based on the reinforcement outcome r_t via a prediction error δ_t:

$$Q_{t+1} = Q_t + \alpha\delta_t$$
$$\delta_t = r_t - Q_t.$$

The reinforcement learning models included two free parameters, α and β. The learning rate α is a parameter between 0 and 1 that measures the extent to which value is updated by feedback from a single trial. Higher α indicates more rapid updating based on few trials and lower α indicates slower updating based on more trials. Another parameter fit to each subject is the inverse temperature parameter β, which determines the probability of making a particular choice using a softmax function (Daw, 2011; Ishii et al., 2002), so that the probability of choosing choice 1 on trial t would be:

$$p(c_t = 1 \mid \alpha, \beta) = \frac{e^{\beta Q_{1t}}}{e^{\beta Q_{1t}} + e^{\beta Q_{2t}}},$$

where $p(c_t = 1)$ refers to the probability of choice one and Q_{1t} is the value for this choice on trial t.

The two free parameters, α and β, were fit to each participant's behavioral data by maximizing an estimate of the log likelihood using a minimizing optimizer (fmincon in MATLAB) with the negative log likelihood as the error function. To constrain the

parameter space to reduce noise, we regularized this estimation with prior distributions on α and β (Daw, 2011):

$$\beta \sim Gamma(1.2, 5)$$
$$\alpha \sim Beta(1.1, 1.1).$$

We also initialized the minimization at 20 random values 2000 times per subject in order to avoid local minima. We selected the optimal α and β parameters for each subject across all trials.

We assessed model fit using Akaike Information Criteria (AIC) to penalize model complexity (Doll et al., 2009; Wagenmakers and Farrell, 2004). We computed the AIC for each subject's likelihood, and subtracted a chance level model fit, then computed a one-way *t*-test on the difference scores.

MRI Acquisition. MRI images were acquired on a 3 T Siemens Tim Trio scanner using a 12-channel head coil. For each block of the learning phase of the conditioning task, we acquired 200 interleaved T2*-weighted echo-planar (EPI) volumes with the following sequence parameters: TR = 2000 ms; TE = 30 ms; flip angle (FA) = 90°; array =64 x 64; 34 slices; effective voxel resolution = 3x3x4 mm; FOV = 192 mm). A high resolution T1-weighted MPRAGE image was acquired for registration purposes (TR = 2170 ms, TE = 4.33 ms, FA = 7º, array = 256 x 256, 160 slices, voxel resolution = 1 mm^3, FOV = 256).

fMRI Preprocessing and Dynamic Connectivity Analysis. Functional images were preprocessed using FSL's FMRI Expert Analysis Tool (FEAT (Smith et al., 2004)). Images from each learning block were high-pass filtered at $f > 0.008$ Hz, spatially smoothed with a 5mm FWHM Gaussian kernel, grand-mean scaled, and motion corrected to their median image using an affine transformation with tri-linear interpolation. The first three images were removed to account for saturation effects. Functional and anatomical images were skull-stripped using FSL's Brain Extraction Tool. Functional images from each block were co-registered to subject's anatomical images and non-linearly transformed to a standard template (T1 Montreal Neurological Institute template, voxel dimensions 2 mm^3) using FNIRT (Andersson et al., 2008). Following image registration, time courses were extracted for each block from 110 cortical and subcortical regions of interest (ROIs) segmented from FSL's Harvard-Oxford Atlas. Due to known effects of motion on measures of functional connectivity (Power et al., 2012; Satterthwaite et al., 2012b), time courses were further preprocessed via a nuisance regression. This regression included the six translation and rotation parameters from the motion correction transformation, average CSF, white matter, and whole brain time courses, as well as the first derivatives, squares, and squared derivatives of each of these confound predictors (Satterthwaite et al., 2012a).

To assess dynamic connectivity between the ROIs, time courses were further subdivided into sub-blocks of 25 TRs each. For each sub-block, connectivity was quantified as the magnitude-squared coherence between each pair of ROIs at $f =$

0.06-0.12 Hz in order to later assess modularity over short time windows in a manner consistent with previous reports (Bassett et al., 2011; Bassett et al., 2013b):

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)} ,$$

where $G_{xy}(f)$ is the cross-spectral density between regions x and y, and $G_{xx}(f)$ and $G_{yy}(f)$ are the autospectral densities of signals x and y, respectively. We thus created subject-specific 110 x 110 x 8 connectivity matrices for 110 regions and 8 time windows for each of the 4 learning blocks, containing coherence values ranging between 0 and 1. The frequency range of 0.06-0.12 Hz was chosen to approximate the frequency envelope of the hemodynamic response, allowing us to detect changes as slow as 3 cycles per window with a 2 second TR.

Network Construction. Each connectivity matrix is treated as a graph or network, in which each brain region is represented as a network node, and each functional connection between two brain regions is represented as a network edge (Bullmore and Sporns, 2009; Bullmore and Bassett, 2011). In the context of dynamic functional connectivity matrices, the network representation is more aptly described as a so-called *temporal network*, which is an ensemble of graphs that are ordered in time (Holme and Saramäki, 2011). If the temporal network contains the same nodes in each graph, then the network is said to be a multilayer network where each layer represents a different time window (Kivelä et al., 2014). The study of topological structure in multilayer networks has been the topic of considerable study in recent years, and many graph metrics and statistics have been extended from the single-network representation to the multilayer network representation. Perhaps one of the single most powerful features of these extensions has been the definition of so-called *identity links*, a new type of edge that links one node in one time slice to itself in the next time slice. These identity links hard code node identity throughout time, and facilitate mathematical extensions and statistical inference in cases that had previously remained challenging.

Uncovering Evolving Circuits Using Multi-slice Community Detection. While many statistics are available to the researcher to characterize network organization in temporal and multilayer networks, it is not entirely clear that all of these statistics are equally valuable in inferring neurophysiologically relevant processes and phenomena (Medaglia et al., 2015). Indeed, many of these statistics are difficult to interpret in the context of neuroimaging data, leading to confusion in the wider literature. A striking contrast to these difficulties lies in the graph-based notion of modularity or community structure (Newman, 2004), which describes the clustering of nodes into densely interconnected groups that are referred to as *modules* or *communities* (Fortunato, 2010; Porter et al., 2009). Recent and convergent evidence demonstrates that these modules can be extracted from rest and task-based fMRI data (Cole et al., 2014; Meunier et al., 2010), demonstrate strong correspondence to known cognitive systems (including default mode, fronto-parietal, cingulo-opercular, visual, auditory, motor, dorsal attention, ventral attention, and subcortical systems (Power et al., 2011; Yeo et al., 2011)), and display

non-trivial re-arrangements during motor skill acquisition (Bassett et al., 2011; Bassett et al., 2015) and memory processing (Braun et al., 2015). These studies support the utility of module-based analyses in the examination of higher order cognitive processes in functional neuroimaging data.

To extract modules or communities from a single-network representation, one typically applies a community detection technique such as modularity maximization (Newman, 2004). However, these single-network algorithms do not allow for the linking of communities across time-points, thus hampering statistically robust inference regarding the reconfiguration of communities as the system evolves (Mucha et al., 2010). In contrast, the multilayer approaches introduced above allow for the characterization of multi-layer network modularity, with layers representing time windows within a learning block. In this framework, each network node (or ROI) in the multi-layer network is connected to itself in the preceding and following time windows in order to link networks in time. This organization enables us to solve the famous community-matching problem explicitly within the model (Mucha et al., 2010), and in principle also facilitates the examination of module reconfiguration across multiple temporal resolutions of system dynamics (Bassett et al., 2013a). We thus constructed multilayer networks for each subject for each of the four learning blocks in the experiment, allowing for the partitioning of each network into communities or modules whose identity is robustly tracked across time windows.

The partitioning of these multilayer networks into temporally linked communities was carried out using a Louvain-like locally greedy algorithm for multilayer modularity optimization (Jutla et al., 2011; Mucha et al., 2010).

$$Q_{ml} = \frac{1}{2\mu}\sum_{ijlr}\left\{\left(A_{ijl} - \gamma_l \frac{k_{il}k_{jl}}{2m_l}\right)\delta_{lr} + \delta_{ij}C_{jlr}\right\}\delta\left(g_{il}, g_{jr}\right),$$

where Q_{ml} is the multilayer modularity index. The adjacency matrix for each layer l consists of components A_{ijl}. The variable γ_l represents the resolution parameter for layer l, while C_{jlr} gives the coupling strength between node j at layers l and r (see below for details of fitting these two parameters). The variables g_{il} and g_{jr} correspond to the community labels for node i at layer l and node j at layer r, respectively; k_{il} is the connection strength (in this case, coherence) of node i in layer l; $2\mu = \sum_{jr}\kappa_{jr}$; the multilayer node strength $\kappa_{jl} = k_{jl} + c_{jl}$; and $c_{jl} = \sum_r C_{jlr}$. Finally, the function $\delta\left(g_{il}, g_{jr}\right)$ refers to the Kronecker delta function, which equals 1 if g_{il}=g_{jr}, and 0 otherwise.

Resolution and coupling parameters (γ_l and C_{jlr}, respectively) were selected using a grid search formulated explicitly to optimize Q_{ml} relative to a temporal null model (Bassett et al., 2013a). The temporal null model we employed is one in which the order of time windows in the multilayer network was permuted uniformly at random. Thus, we performed a grid search to identify the values of γ_l and C_{jlr} that maximized $Q_{ml} - Q_{null}$, following (Bassett et al., 2013a). To ensure statistical

robustness, we repeated this grid search 10 times. To maximize the stability of resolution and coupling, each subject's parameters were treated as random effects, with the best estimate of resolution and coupling generated by averaging across-individual subject estimates. This is a similar approach to that taken in computational modeling of reinforcement learning, in which learning rate and temperature parameters are averaged in order to generate prediction error estimates(Daw, 2011). With this approach, we estimated the optimal resolution parameter γ to be 1.18 and the coupling parameter C to be 1. These values are quite similar to those chosen *a priori* (usually setting both parameters to unity) in previous reports (Bassett et al., 2011).

Finally, we note that maximization of the modularity quality function is NP-hard, and the Louvain-like locally greedy algorithm we employ is a computational heuristic with non-deterministic solutions. Due to the well known near-degeneracy of Q_{ml} (Bassett et al., 2013a; Good et al., 2010; Mucha et al., 2010), we repeated the multi-slice community detection algorithm 500 times using the resolution and coupling parameters estimated from the grid search procedure outlined above. This approach ensured an adequate sampling of the null distribution (Bassett et al., 2013a). Each repetition produced a hard partition of nodes into communities as a function of time window: that is, a community allegiance identity for each of the 110 brain regions in the multilayer network.

Flexibility –A Dynamic Network Statistic. To characterize the dynamics of these temporal networks and their relation to learning, we computed the *flexibility* of each node, which measures the extent to which a region changed its community allegiance over time (Bassett et al., 2011) . Intuitively and more colloquially, flexibility can be thought of as a measure of a region's tendency to communicate with different networks during learning. Flexibility for a learning block is defined as the number of times a node displays a change in community assignment over time, divided by the number of possible changes (which is equal to the number of time points in that block minus 1). To obtain a stable estimate of flexibility, we averaged flexibility scores for each ROI over the 500 iterations of the multilayer community detection algorithm described above. This measure was computed for each region in each block. In addition, average measures of flexibility were computed across the whole-brain and across all blocks. To assess the spatial distribution of flexibility over the brain, we visualized the average flexibility over all blocks and subjects for each region using surface maps.

Relating Flexibility to Reinforcement Learning. To examine the effect of flexibility on learning from feedback, we estimated a generalized mixed-effects model predicting optimally correct choices with flexibility estimates for each block with a logistic link function, using the Maximum Likelihood (ML) approximation implemented in the *lme4* package (Bates et al., 2015). Subjects' average flexibility in an *a priori* striatum ROI was used to predict the proportion of optimal choices in each learning block. The ROI included bilateral caudate, putamen, and nucleus accumbens regions from the Harvard-Oxford atlas. We included a random effect of subject, allowing for

different effects of flexibility on learning for each subject, while constraining these effects with the group average. Average flexibility across sessions was included as a fixed effect in the model in order to ensure that our estimates represented within-subject learning effects. We also estimated this effect for whole-brain flexibility, which has related to several cognitive functions in previous reports (Bassett et al., 2011; Braun et al., 2015).

To provide appropriate posterior inference about the plausible parameter values indicated by our data, and to account for uncertainty about all parameters, we also fit a fully Bayesian extension of the ML approximation described above for the effect of striatal flexibility on learning performance (**Supplemental Figure 1**). We used the 'brms' package for fitting models in the Stan language (Carpenter et al., 2015). These models were similar to the likelihood approximation models, but included a covariance parameter for subject-level slopes and intercepts (which could not be fit by the above approximation), and weakly informative prior distributions to regularize parameter estimation:

$$\beta \sim N(0,10^2)$$

$$\tau \sim T^+(3, 0, 5),$$

where β represents the "fixed effects" parameters (slope and intercept), τ represents the "random effects" variance for subject-level estimates sampled from β, and T^+ is a positive half-t distribution (Gelman, 2006). We used an lkj(2) prior for correlations between subject-level intercept and slope estimates (Lewandowski et al., 2009).

To explore other regions exhibiting effects of dynamic connectivity on learning performance, we separately modeled the effect of flexibility in each brain region on reinforcement learning using the ML approximation implemented in the *lme4* package. To control for Type I error, we applied a false discovery rate correction for multiple comparisons across regions (Benjamini and Hochberg, 1995). While regions passing this threshold are reported, we also visualize the results using an exploratory uncorrected threshold of $p<0.05$ (**Supplemental Figure 3**).

To examine the relationship between flexibility and parameters estimated from standard reinforcement learning models, we tested whether whole-brain and striatal flexibility were correlated with the learning rate α for each subject, using Spearman's correlation coefficient due to the non-Gaussian distribution of learning rates. If flexibility is related to the integration of sensory, motor, and value information over time, it should be associated with learning rate, which measures the extent to which subjects average feedback from individual trials when updating instrumental value, as opposed to averaging over many experiences. Lower values of the learning rate indicate integration of information across more trials; higher values of learning rate indicate integration of information across fewer trials.

To explore the relationship between network dynamics and other forms of learning, we also regressed flexibility statistics from all ROIs against subsequent memory

scores for the trial-unique objects presented during feedback. If the effects of striatal flexibility are relatively selective to incremental learning, we expected to find no significant association even at an uncorrected threshold with memory in the regions comprising our striatal ROI. In addition, this provided an exploratory analysis to examine the regions in which network flexibility plays a potential role in episodic memory. Given a host of previous studies on multiple learning systems, we reasoned it might be possible to detect an effect of dynamic network coupling on episodic memory in regions traditionally associated with this form of learning.

Acknowledgements: DSB acknowledges support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Army Research Laboratory and the Army Research Office through contract numbers W911NF-10-2-0022 and W911NF-14-1-0679, the National Institute of Mental Health (2-R01-DC-009209-11), the National Institute of Child Health and Human Development (1R01HD086888-01), the Office of Naval Research, and the National Science Foundation (CRCNS award BCS-1441502 and CAREER award PHY-1554488). AG acknowledges support from the William T. Grant Foundation and the Center for Translational and Prevention Science (P30 DA027827, Brody-PI) funded by NIDA. D.S. acknowledges support from the National Science Foundation (Career Award #0955494) and the National Institute of Health (NINDS grant # R01NS079784 and CRCNS grant #R01DA038891).

Figures

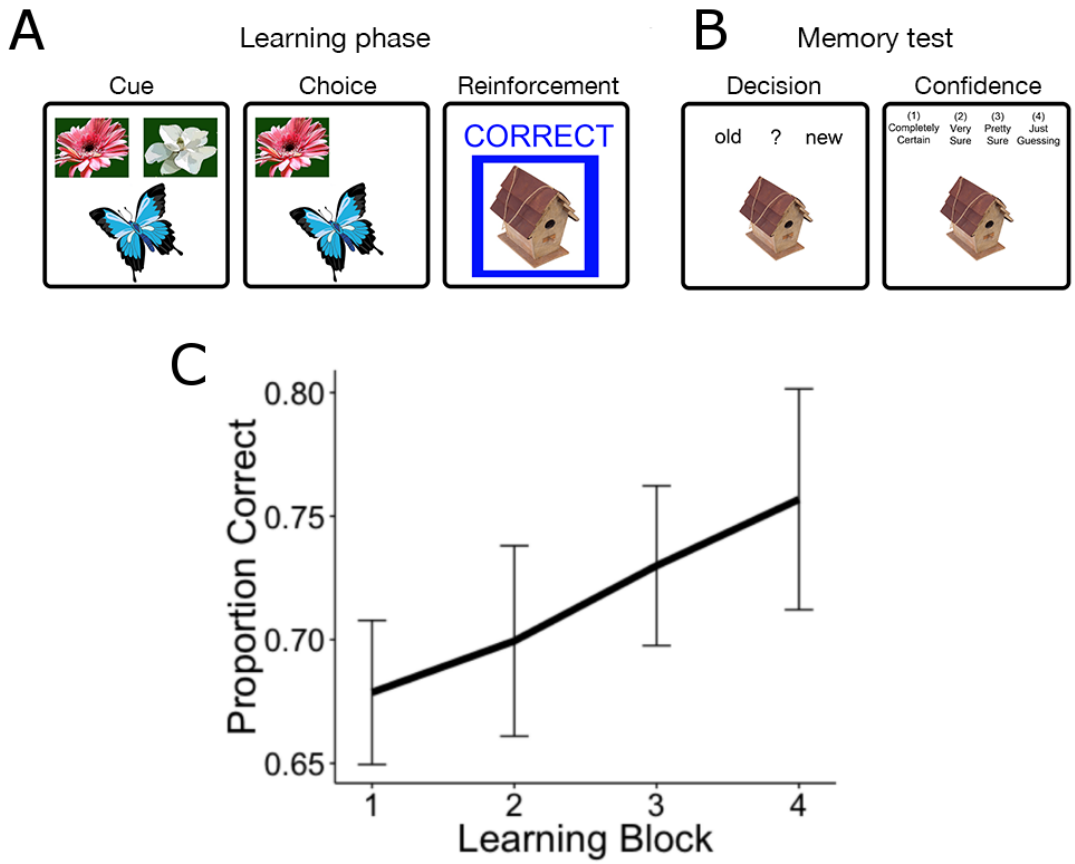


Figure 1. Task design and learning performance. Participants performed a modified reinforcement learning task while undergoing fMRI(Foerde and Shohamy, 2011). **A.** Participants were instructed to associate each of 4 cues (butterflies) with one of two outcomes (flowers). Feedback was probabilistic, with positive feedback following the choice on 80% of correct trials and on 20% of incorrect trials. **B.** Each feedback event was presented with a unique image. Thirty minutes following the MRI scan, participants were given a surprise episodic memory test, testing recognition and confidence for images seen during the scan, intermixed with novel images. **C.** Average performance on the learning task improved linearly, suggesting continuous learning across all trials.

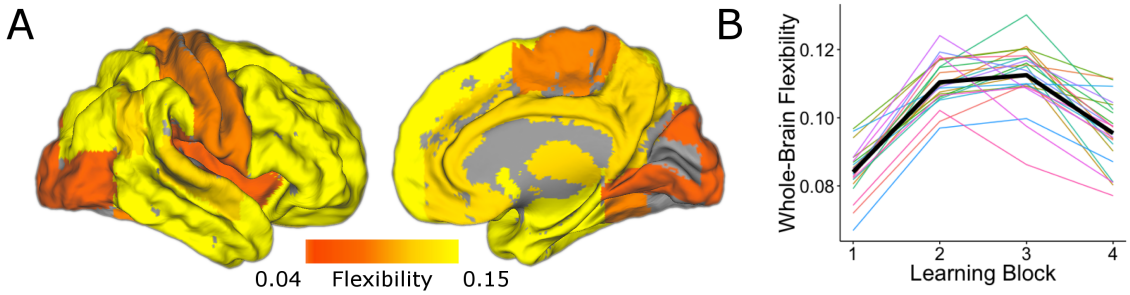


Figure 2. Spatial and temporal characteristics of network flexibility. Network flexibility exhibited distinct spatial and temporal patterns. **A.** Flexibility averaged across all learning blocks was highest across regions of association cortex and lowest in sensory and motor regions. **B.** Flexibility averaged across the whole brain consistently increased in early learning blocks. Color lines represent whole-brain flexibility for single subjects.

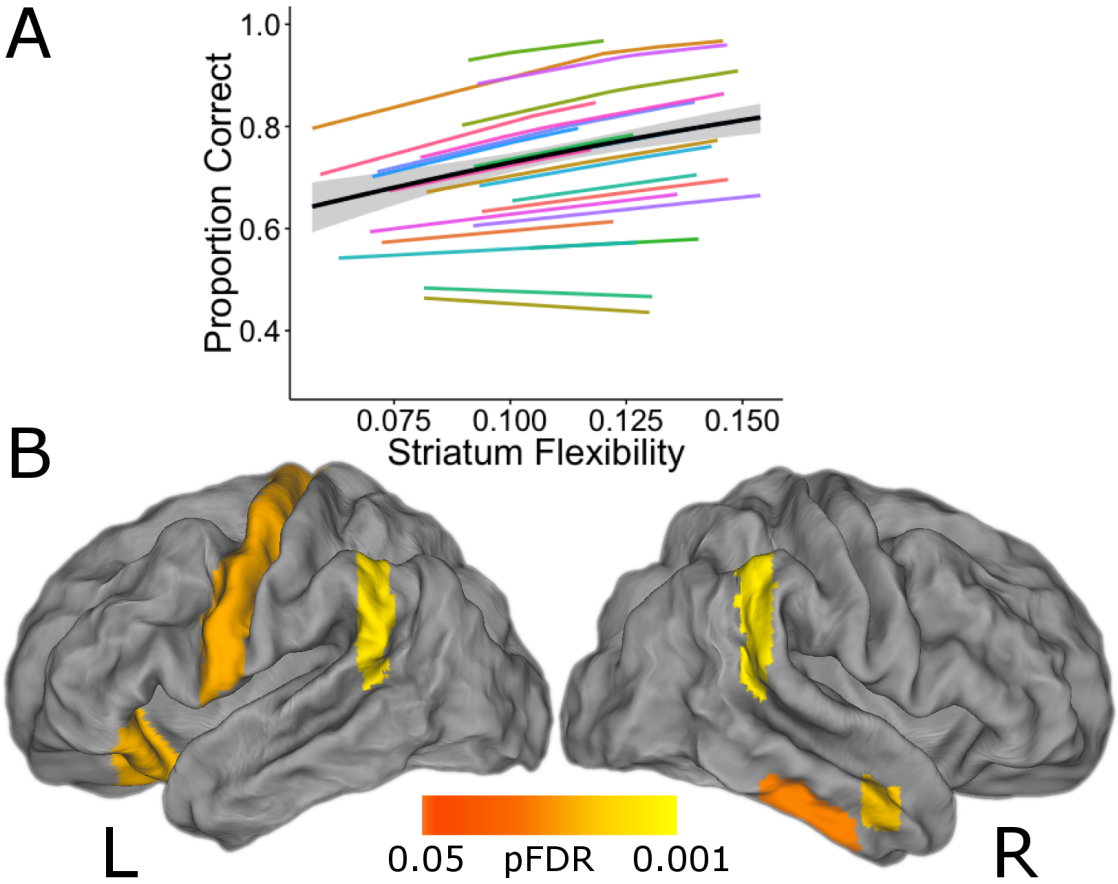


Figure 3. Flexibility in the striatum and beyond relates to learning performance. Network flexibility in a number of brain regions is related to learning performance within subjects. **A.** Mixed-effects model fit for the association between network flexibility in an *a priori* striatum ROI and learning performance. The black line represents the fixed effect estimate and the gray band represents the 95% confidence interval for this estimate, while color lines represent subject-level random effects estimates. See **Supplemental Information** for a Bayesian extension of this model and full model fits with uncertainty for individual subjects. **B.** Regions passing FDR correction following a univariate whole-brain analysis using the same model as above. Regions passing this threshold include left motor cortex, bilateral parietal cortex, and right orbitofrontal cortex. See **Supplemental Information** for a full list of regions and an exploratory uncorrected map.

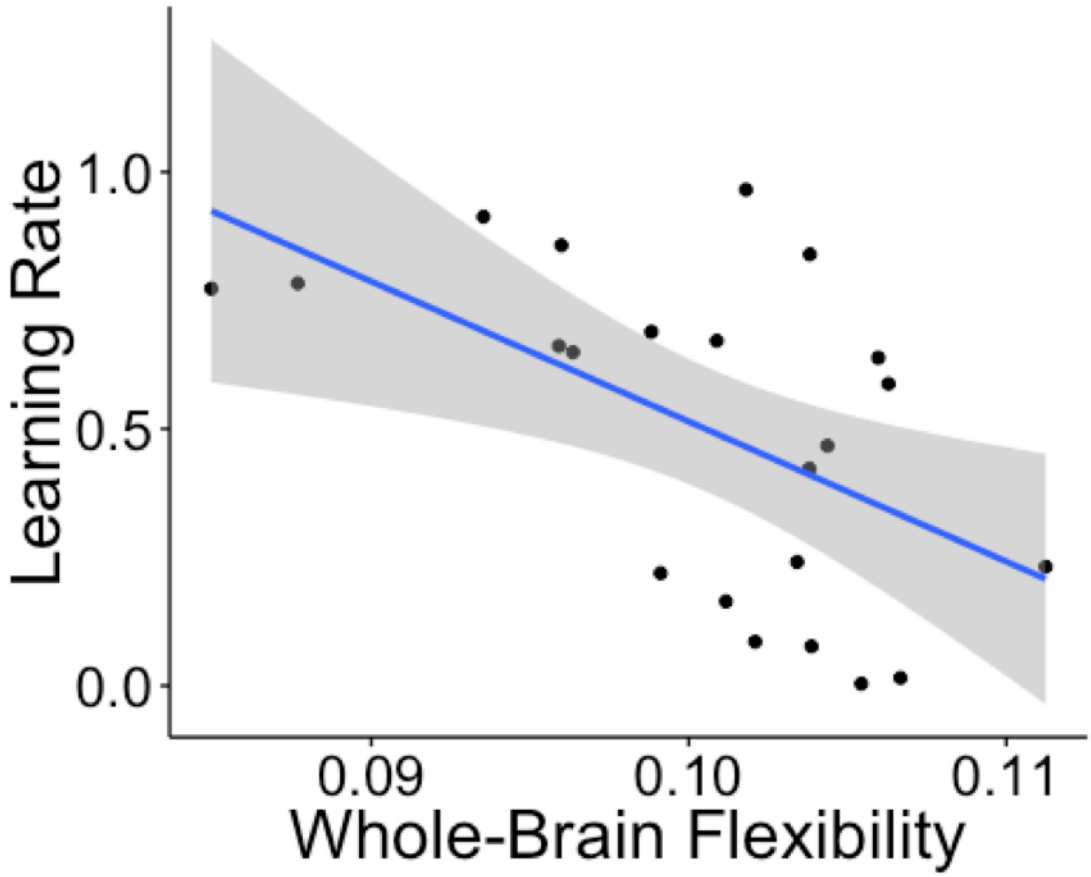


Figure 4. Whole-brain flexibility is negatively correlated with single-trial learning rate. Standard reinforcement learning models were fit to each participant's choice behavior. The learning rate parameter from these models, α, indexes subjects' tendency to weight single trials (high α) as opposed to averaging over many experiences (low α), and was negatively correlated with whole-brain flexibility. These results indicate that whole-brain flexibility is greater when individuals are integrating information across many trials. Grey band represents 95% confidence interval.

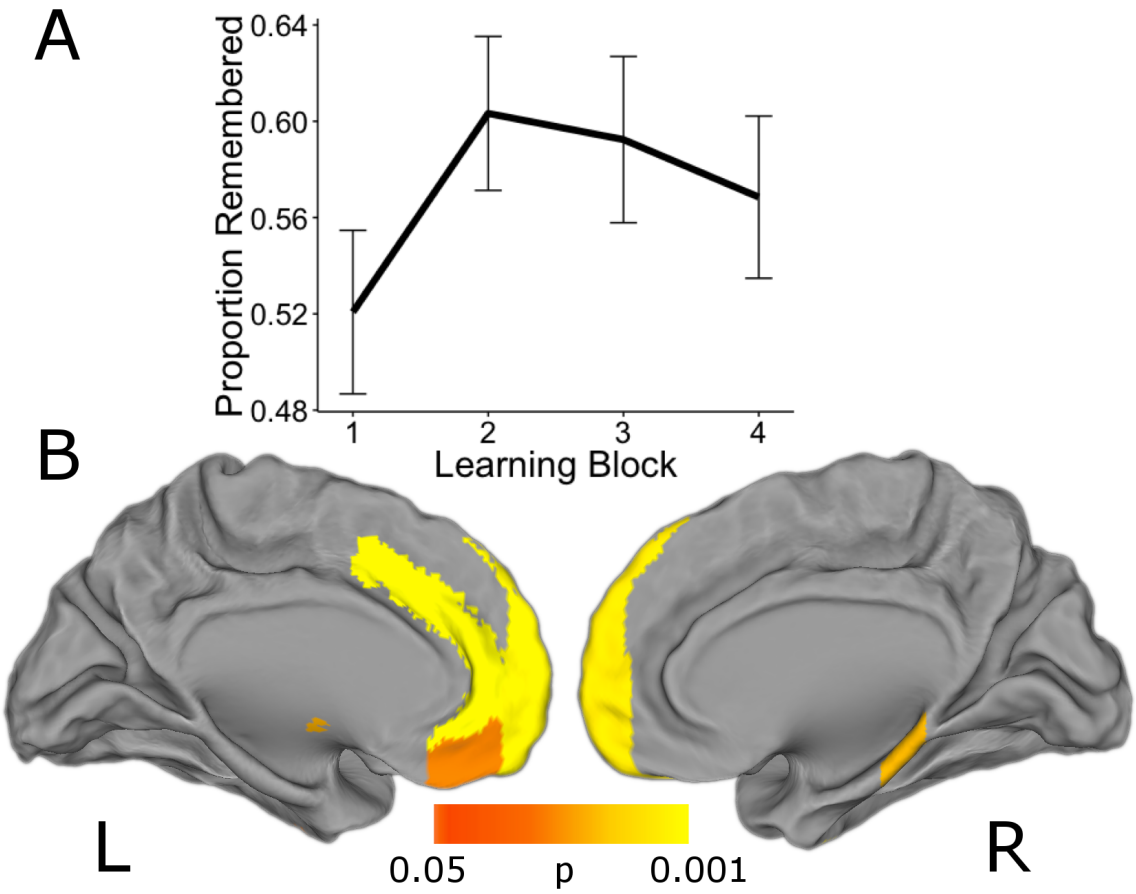


Figure 5. Network flexibility in medial prefrontal cortex relates to episodic memory. An exploratory analysis showed effects of network flexibility on episodic memory performance in medial prefrontal and temporal lobes. **A.** Average memory (% correct) across trials. Participants' recollection accuracy varied across blocks. Line represents group average and bars represent standard errors. **B.** A number of medial prefrontal regions as well as the right parahippocampal gyrus passed an exploratory uncorrected threshold of $p<0.05$ for the effect of flexibility on subsequent episodic memory. The effect in the left paracingulate gyrus survived FDR correction.

References

Alexander, G.E., DeLong, M.R., and Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual review of neuroscience 9, 357-381.

Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., and Calhoun, V.D. (2012). Tracking whole-brain connectivity dynamics in the resting state. Cerebral cortex, bhs352.

Andersson, J., Smith, S., and Jenkinson, M. (2008). Fnirt-fmrib's non-linear image registration tool. Human Brain Mapping 2008.

Bassett, D.S., Porter, M.A., Wymbs, N.F., Grafton, S.T., and Carlson, J.M. (2013a). Robust detection of dynamic community structure in networks. Chaos 23, 013142.

Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., and Grafton, S.T. (2011). Dynamic reconfiguration of human brain networks during learning. Proceedings of the National Academy of Sciences 108, 7641-7646.

Bassett, D.S., Wymbs, N.F., Rombach, M.P., Porter, M.A., Mucha, P.J., and Grafton, S.T. (2013b). Task-based core-periphery organization of human brain dynamics. PLoS computational biology 9, e1003171.

Bassett, D.S., Yang, M., Wymbs, N.F., and Grafton, S.T. (2015). Learning-induced autonomy of sensorimotor systems. Nature Neuroscience.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. Journal of Statistical Software 67.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological), 289-300.

Betzel, R.F., Satterthwaite, T.D., Gold, J.I., and Bassett, D.S. (2016). A positive mood, a flexible brain. arXiv preprint arXiv:160107881.

Bogacz, R., and Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. Neural computation 19, 442-477.

Braun, U., Schäfer, A., Bassett, D.S., Rausch, F., Schweiger, J.I., Bilek, E., Erk, S., Romanczuk-Seiferth, N., Grimm, O., and Geiger, L.S. (2016). Dynamic brain network reconfiguration as a potential schizophrenia genetic risk mechanism modulated by NMDA receptor function. Proceedings of the National Academy of Sciences 113, 12568-12573.

Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., Schweiger, J.I., Grimm, O., Heinz, A., and Tost, H. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. Proceedings of the National Academy of Sciences 112, 11678-11683.

Bressler, S.L., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. Trends in cognitive sciences 14, 277-290.

Büchel, C., Coull, J., and Friston, K. (1999). The predictive value of changes in effective connectivity for human learning. Science 283, 1538-1541.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10, 186-198.

Bullmore, E.T., and Bassett, D.S. (2011). Brain graphs: graphical models of the human brain connectome. Annual review of clinical psychology 7, 113-140.

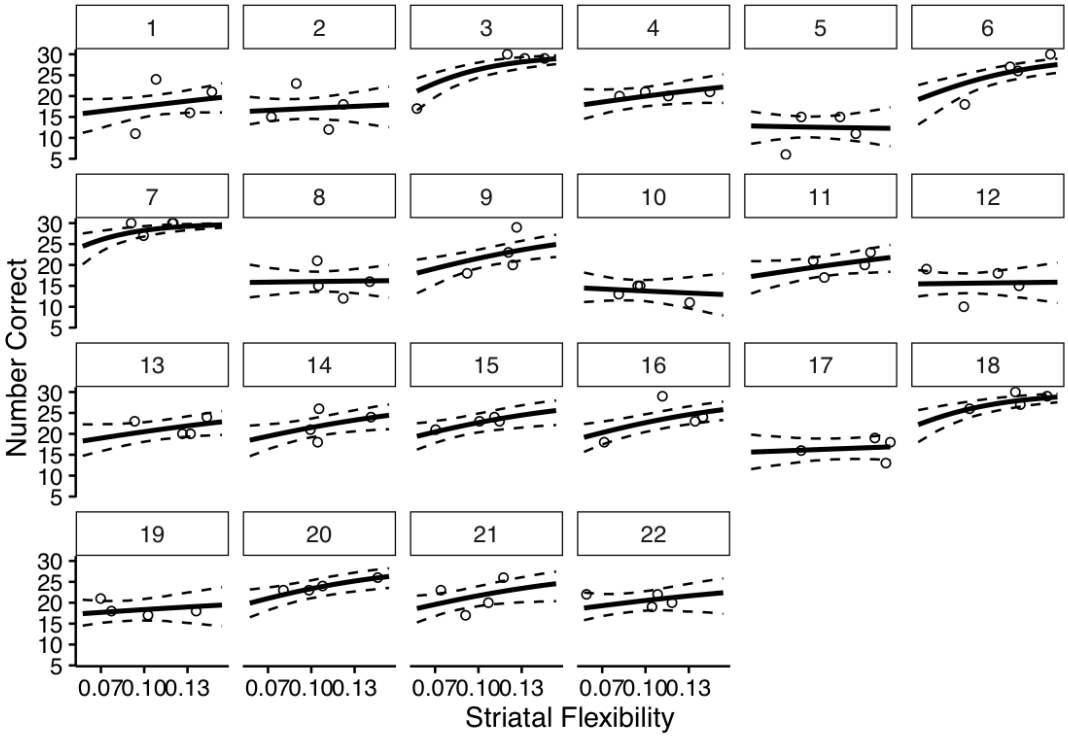
Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., and Riddell, A. (2015). Stan: a probabilistic programming language. Journal of Statistical Software.
Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., and Petersen, S.E. (2014). Intrinsic and task-evoked network architectures of the human brain. Neuron 83, 238-251.
Damaraju, E., Allen, E., Belger, A., Ford, J., McEwen, S., Mathalon, D., Mueller, B., Pearlson, G., Potkin, S., and Preda, A. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage: Clinical 5, 298-308.
Damoiseaux, J., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C., Smith, S.M., and Beckmann, C. (2006). Consistent resting-state networks across healthy subjects. Proceedings of the national academy of sciences 103, 13848-13853.
Davidow, J.Y., Foerde, K., Galván, A., and Shohamy, D. (2016). An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. Neuron 92, 93-99.
Daw, N.D. (2011). Trial-by-trial data analysis using computational models. Decision making, affect, and learning: Attention and performance XXIII 23, 3-38.
Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience 8, 1704-1711.
Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. Nature 441, 876-879.
Ding, L. (2015). Distinct dynamics of ramping activity in the frontal cortex and caudate nucleus in monkeys. Journal of neurophysiology 114, 1850-1861.
Doll, B.B., Jacobs, W.J., Sanfey, A.G., and Frank, M.J. (2009). Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. Brain research 1299, 74-94.
Durante, D., Mukherjee, N., and Steorts, R.C. (2016). Bayesian Learning of Dynamic Multilayer Networks. arXiv preprint arXiv:160802209.
Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. Nature Reviews Neuroscience 1, 41-50.
Fineberg, N.A., Potenza, M.N., Chamberlain, S.R., Berlin, H.A., Menzies, L., Bechara, A., Sahakian, B.J., Robbins, T.W., Bullmore, E.T., and Hollander, E. (2010). Probing compulsive and impulsive behaviors, from animal models to endophenotypes: a narrative review. Neuropsychopharmacology 35, 591-604.
Foerde, K., Braun, E.K., and Shohamy, D. (2012). A trade-off between feedback-based learning and episodic memory for feedback events: evidence from Parkinson's disease. Neurodegenerative Diseases 11, 93-101.
Foerde, K., Race, E., Verfaellie, M., and Shohamy, D. (2013). A role for the medial temporal lobe in feedback-driven learning: evidence from amnesia. The Journal of Neuroscience 33, 5698-5704.
Foerde, K., and Shohamy, D. (2011). Feedback timing modulates brain systems for learning in humans. The Journal of Neuroscience 31, 13157-13167.
Foerde, K., Steinglass, J.E., Shohamy, D., and Walsh, B.T. (2015). Neural mechanisms supporting maladaptive food choices in anorexia nervosa. Nature neuroscience.
Fortunato, S. (2010). Community detection in graphs. Physics reports 486, 75-174.

Frank, M.J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T.V., Cavanagh, J.F., and Badre, D. (2015). fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. The Journal of Neuroscience 35, 485-494.
Frank, M.J., Seeberger, L.C., and O'reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940-1943.
Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian analysis 1, 515-534.
Gerraty, R.T., Davidow, J.Y., Wimmer, G.E., Kahn, I., and Shohamy, D. (2014). Transfer of learning relates to intrinsic connectivity between hippocampus, ventromedial prefrontal cortex, and large-scale networks. The Journal of Neuroscience 34, 11297-11303.
Good, B.H., de Montjoye, Y.-A., and Clauset, A. (2010). Performance of modularity maximization in practical contexts. Physical Review E 81, 046106.
Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. Current opinion in neurology 21, 424-430.
Haber, S.N. (2003). The primate basal ganglia: parallel and integrative networks. Journal of chemical neuroanatomy 26, 317-330.
Haber, S.N., Kim, K.-S., Mailly, P., and Calzavara, R. (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. The Journal of Neuroscience 26, 8368-8376.
Haber, S.N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35, 4-26.
Harmelech, T., Preminger, S., Wertman, E., and Malach, R. (2013). The day-after effect: long term, Hebbian-like restructuring of resting-state fMRI patterns induced by a single epoch of cortical activation. The Journal of Neuroscience 33, 9488-9497.
Hikosaka, O., Kim, H.F., Yasuda, M., and Yamamoto, S. (2014). Basal ganglia circuits for reward value-guided behavior. Annual review of neuroscience 37, 289.
Holme, P., and Saramäki, J. (2011). Temporal Networks. Physics Reports 519, 97-125.
Ishii, S., Yoshida, W., and Yoshimoto, J. (2002). Control of exploitation–exploration meta-parameter in reinforcement learning. Neural networks 15, 665-687.
Jutla, I.S., Jeub, L.G., and Mucha, P.J. (2011). A generalized Louvain method for community detection implemented in MATLAB. URL http://netwiki amath unc edu/GenLouvain.
Kehoe, E.J., and Gormezano, I. (1980). Configuration and combination laws in conditioning with compound stimuli. Psychological Bulletin 87, 351.
Kemp, J.M., and Powell, T. (1971). The connexions of the striatum and globus pallidus: synthesis and speculation. Philosophical Transactions of the Royal Society of London B: Biological Sciences 262, 441-457.
Khambhati, A.N., Davis, K.A., Lucas, T.H., Litt, B., and Bassett, D.S. (2016). Virtual cortical resection reveals push-pull network control preceding seizure evolution. Neuron 91, 1170-1182.
Kheirbek, M.A., Klemenhagen, K.C., Sahay, A., and Hen, R. (2012). Neurogenesis and generalization: a new approach to stratify and treat anxiety disorders. Nature neuroscience 15, 1613-1620.

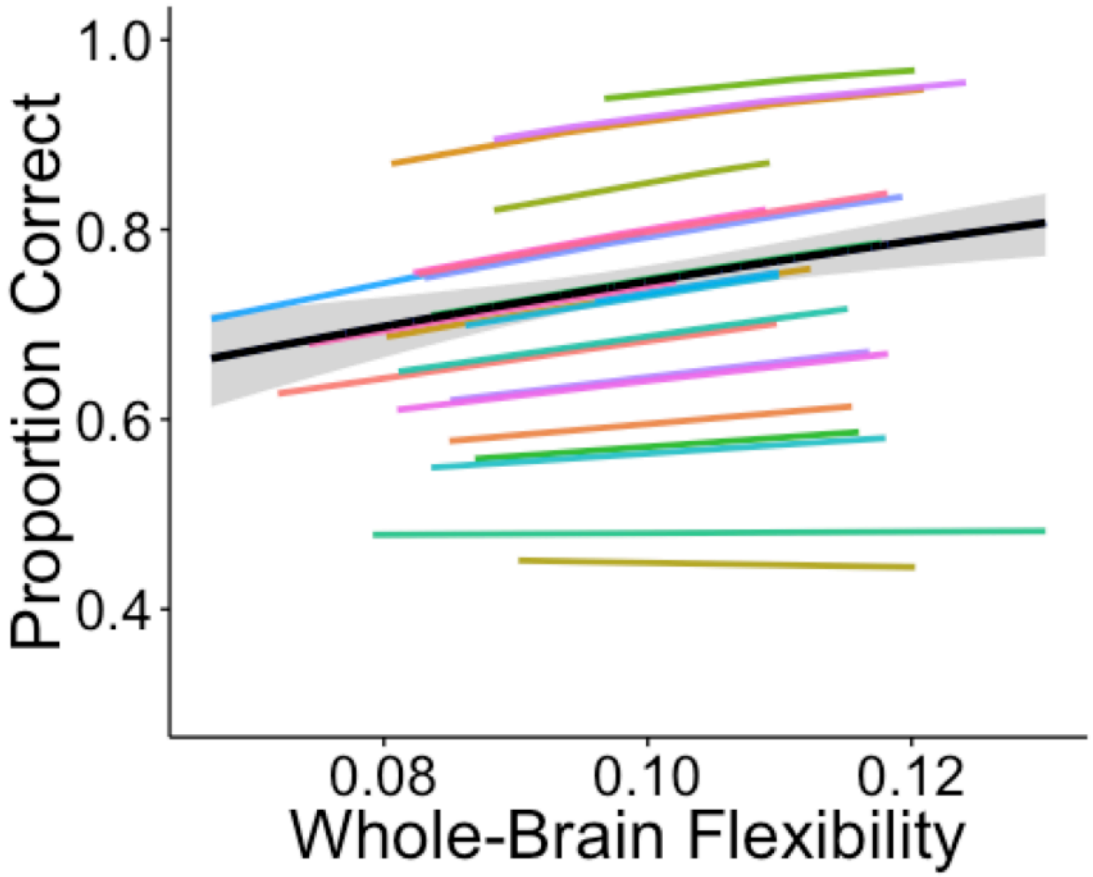
Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., and Porter, M.A. (2014). Multilayer networks. Journal of Complex Networks 2, 203 - 271.
Kopell, N.J., Gritton, H.J., Whittington, M.A., and Kramer, M.A. (2014). Beyond the connectome: the dynome. Neuron 83, 1319-1328.
Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. Journal of multivariate analysis 100, 1989-2001.
Liu, X., and Duyn, J.H. (2013). Time-varying functional network information extracted from brief instances of spontaneous brain activity. Proceedings of the National Academy of Sciences 110, 4392-4397.
Ma, S., Calhoun, V.D., Phlypo, R., and Adalı, T. (2014). Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis. NeuroImage 90, 196-206.
Medaglia, J.D., Lynall, M.-E., and Bassett, D.S. (2015). Cognitive Network Neuroscience. Journal of cognitive neuroscience.
Meunier, D., Lambiotte, R., and Bullmore, E.T. (2010). Modular and hierarchically modular organization of brain networks. Frontiers in neuroscience 4, 200.
Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., and Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. Science 328, 876-878.
Newman, M.E. (2004). Fast algorithm for detecting community structure in networks. Physical review E 69, 066133.
O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal difference models and reward-related learning in the human brain. Neuron 38, 329-337.
Palla, K., Caron, F., and Teh, Y.W. (2016). Bayesian nonparametrics for Sparse Dynamic Networks. arXiv preprint arXiv:160701624.
Porter, M.A., Onnela, J.-P., and Mucha, P.J. (2009). Communities in networks. Notices of the AMS 56, 1082-1097.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142-2154.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., and Schlaggar, B.L. (2011). Functional network organization of the human brain. Neuron 72, 665-678.
Robbins, T.W., Gillan, C.M., Smith, D.G., de Wit, S., and Ersche, K.D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. Trends in cognitive sciences 16, 81-91.
Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., and Gur, R.E. (2012a). An Improved Framework for Confound Regression and Filtering for Control of Motion Artifact in the Preprocessing of Resting-State Functional Connectivity Data. NeuroImage.
Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., and Gur, R.E. (2012b). Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science 275, 1593-1599.
Shine, J.M., Bissett, P.G., Bell, P.T., Koyejo, O., Balsters, J.H., Gorgolewski, K.J., Moodie, C.A., and Poldrack, R.A. (2016). The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. Neuron 92, 544-554.
Shohamy, D., Myers, C., Grossman, S., Sage, J., Gluck, M., and Poldrack, R. (2004). Cortico‐striatal contributions to feedback‐based learning: converging data from neuroimaging and neuropsychology. Brain 127, 851-859.
Siebenhühner, F., Weiss, S.A., Coppola, R., Weinberger, D.R., and Bassett, D.S. (2013). Intra-and inter-frequency brain network structure in health and schizophrenia. PloS one 8, e72351.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., and Flitney, D.E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208-S219.
Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann, C.F., Jenkinson, M., Andersson, J., and Glasser, M.F. (2012). Temporally-independent functional modes of spontaneous brain activity. Proceedings of the National Academy of Sciences 109, 3131-3136.
Staresina, B.P., and Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. Neuron 63, 267-276.
Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: An introduction (MIT press).
Tambini, A., Ketz, N., and Davachi, L. (2010). Enhanced brain correlations during rest are related to memory for recent experiences. Neuron 65, 280-290.
Wagenmakers, E.-J., and Farrell, S. (2004). AIC model selection using Akaike weights. Psychonomic bulletin & review 11, 192-196.
Waltz, J.A., and Gold, J.M. (2007). Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. Schizophrenia research 93, 296-303.
Wiecki, T.V., and Frank, M.J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. Psychological Review 120, 329.
Wimmer, G.E., Daw, N.D., and Shohamy, D. (2012). Generalization of value in reinforcement learning by humans. European Journal of Neuroscience 35, 1092-1104.
Wimmer, G.E., and Shohamy, D. (2012). Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. Science 338, 270-273.
Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., and Polimeni, J.R. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology 106, 1125-1165.

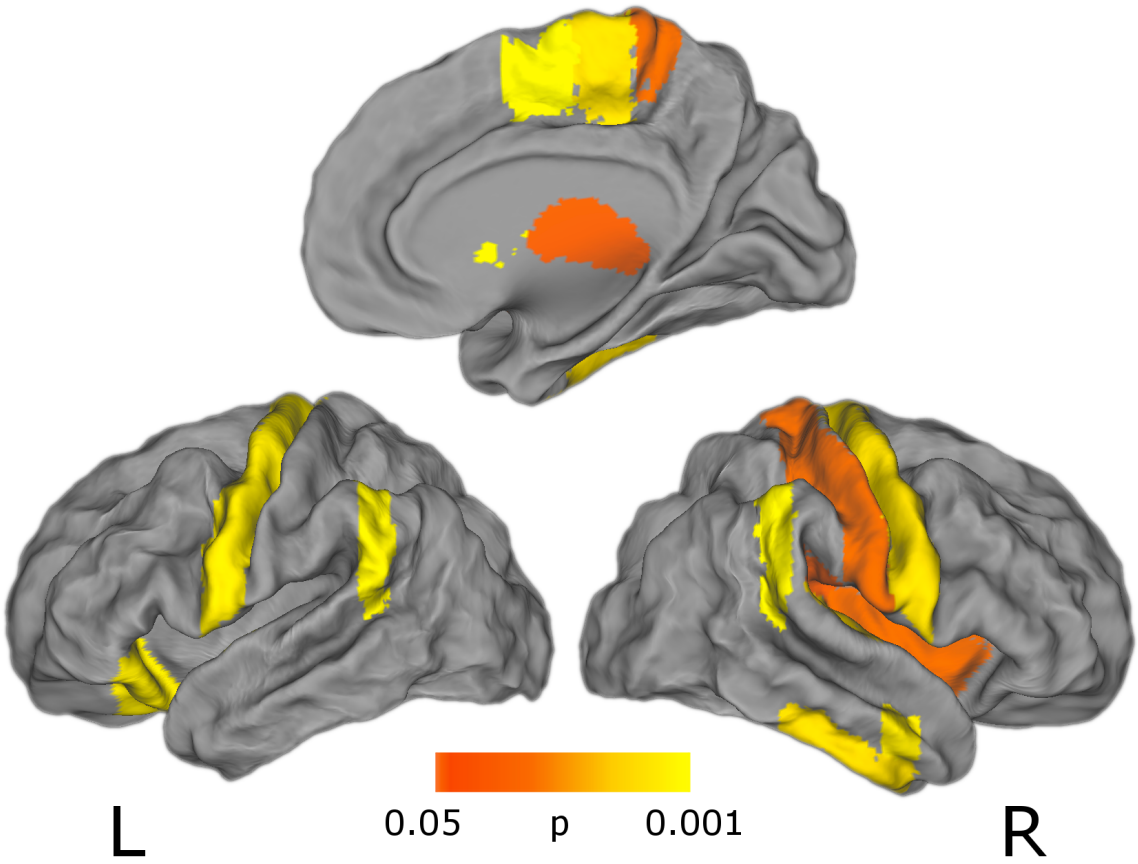
Supplemental Figures



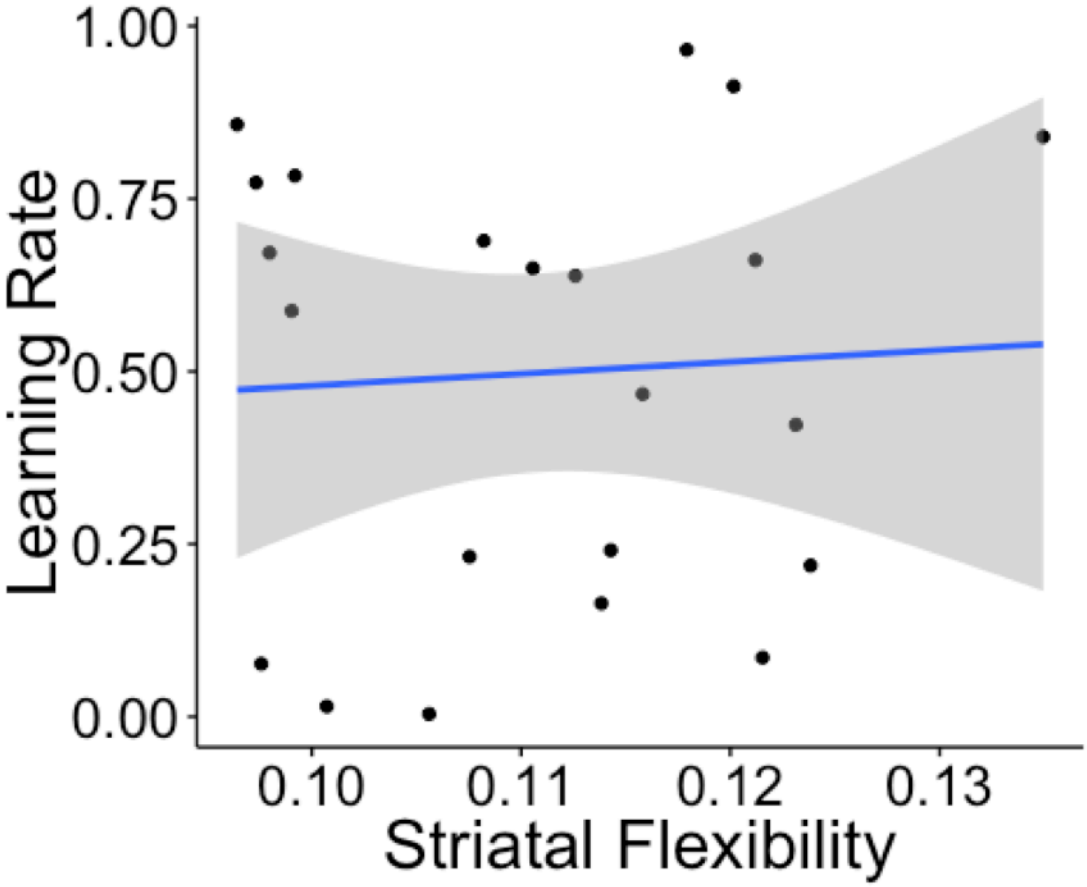
Supplemental Figure 1. Subject-level data and fits for Bayesian hierarchical model of the effect of striatal flexibility on learning. For posterior inference on the effect of striatal flexibility on learning performance, we fit a Bayesian hierarchical model. Each subplot displays data (open circles) from a single subject. Solid lines represent model estimates for the effect of flexibility on learning, while dotted lines represent 95% credible intervals.



Supplemental Figure 2. Whole-brain network flexibility is related to learning performance. Given the distributed effect of flexibility on reinforcement learning, and following other work on motor skills and working memory, we fit a mixed effects model predicting learning performance with a global measure of flexibility averaged across the whole brain. The black line represents the fixed effect estimate and the gray band represents the 95% confidence interval for this estimate, while color lines represent subject-level random effects estimates.



Supplemental Figure 3. Exploratory uncorrected ($p<0.05$) results for whole-brain effect of network flexibility on reinforcement learning performance.



Supplemental Figure 4. We found no relationship between flexibility in the striatum and single-trial updating rates derived from reinforcement learning models. Grey band represents 95% confidence interval.

1032 **Table 1**

Harvard Oxford Region	Regression Coefficient	p 1033
Left Caudate	5.93	0.003
Left Orbitofrontal Cortex	4.88	0.006
Left Planum Polare	10.97	0.000
Left Precentral Gyrus	4.79	0.006
Left Supramarginal Gyrus	5.38	0.003
Right Inferior Temporal Gyrus, anterior	4.75	0.01
Right Inferior Temporal Gyrus, posterior	4.66	0.01
Right Supplemental Motor Cortex	8.54	0.0009
Right Middle Temporal Gyrus	3.51	0.006
Right Planum Temporal	8.19	0.01
Right Putamen	6.10	0.003
Right Supramarginal Gyrus	6.48	0.003

1034

1035