

1 Dynamic flexibility in striatal-cortical circuits supports reinforcement
2 learning

3
4 Raphael T. Gerraty¹, Juliet Y. Davidow², Karin Foerde³, Adriana Galvan⁴, Danielle S. Bassett^{5,6},
5 and Daphna Shohamy^{1,7}

6
7 ¹Department of Psychology, Columbia University, New York, NY, USA.

8 ²Department of Psychology, Harvard University, Cambridge MA, USA.

9 ³Department of Psychology, New York University, New York, NY, USA.

10 ⁴Department of Psychology, UCLA, Los Angeles, CA, USA

11 ⁵Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104 USA

12 ⁶Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia,
13 PA 19104 USA

14 ⁷Zuckerman Mind Brain Behavior Institute and Kavli Institute for Brain Science, Columbia
15 University, New York, NY, USA.

16
17 **Summary**

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

33

34

35

36

37

38

39

40

41 **Introduction**

42 Learning from reinforcement is central to adaptive behavior and requires
43 continuous and dynamic integration of sensory, motor, and cognitive information
44 over time. Major progress has been made in understanding how individual brain
45 regions support reinforcement learning. However, remarkably little is known about
46 how these brain regions interact during learning, how their interaction changes
47 over time, and how these dynamic circuit level changes relate to successful learning.
48

49 In a typical reinforcement learning task, participants use trial-by-trial reinforcement
50 over hundreds of trials to learn to associate cues or actions with their most probable
51 outcome (e.g. Daw, 2011; Frank et al., 2004; O'Doherty et al., 2003; Shohamy et al.,
52 2004). Computationally, this sort of learning is captured by so-called "model-free"
53 reinforcement learning algorithms, a class of models that provide a quantitative and
54 mechanistic framework for describing behavior on a trial-by-trial basis (Daw, 2011;
55 Daw et al., 2005; Sutton and Barto, 1998). Such models have also been highly
56 successful in accounting for neuronal signals underlying learning behavior (Daw et
57 al., 2006; O'Doherty et al., 2003; Schultz et al., 1997), collectively demonstrating a
58 central role for the striatum and its dopaminergic inputs in model-free
59 reinforcement learning. To accomplish such learning, however, the striatum and
60 other regions must integrate visual, motor, and reinforcement information over
61 time, a process that is likely to involve coordination across a number of different
62 circuits.

63

64 The idea that the striatum serves such an integrative role in learning and cognition
65 is not new (Bogacz and Gurney, 2007; Ding, 2015; Frank et al., 2015; Haber, 2003;
66 Hikosaka et al., 2014; Kemp and Powell, 1971; Wiecki and Frank, 2013).

67 Anatomically, the striatum is well positioned for such integration, given that it
68 receives extensive input from many regions of cortex and projects back, through
69 thalamus, to motor cortex (Alexander et al., 1986; Haber, 2003; Haber et al., 2006;
70 Haber and Knutson, 2010). However, while the idea that the striatum serves such a
71 role is anatomically and theoretically appealing, it has been very difficult to test this
72 possibility empirically. Thus, whether the striatum interacts with other sensory,
73 motor and cognitive regions during learning, and how such network-level
74 interactions reconfigure over the course of learning, remains unknown.

75

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

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

88

89 Here we aimed to address this gap. We take advantage of recent advances in a
90 dynamic formulation of graph theory and its application to neuroimaging data, an
91 emerging field known as dynamic network neuroscience (Kopell et al., 2014;
92 Medaglia et al., 2015). This formulation has been spurred by the development of
93 algorithmic tools like multi-slice community detection (Mucha et al., 2010), which
94 can be used to infer activated circuits and their reconfiguration from fMRI data
95 collected as participants perform cognitively demanding tasks (Bassett et al., 2011;
96 Bassett et al., 2015; Braun et al., 2015). Such tools have recently been leveraged to
97 understand the role of dynamic brain-wide connectivity in motor skill learning
98 (Bassett et al., 2011; Bassett et al., 2013b; Bassett et al., 2015). A key measure from
99 this emerging field is an index of a brain region's tendency to communicate with
100 different networks over time, known as "flexibility" (Bassett et al., 2011; Braun et al.,
101 2015). Prior work has shown that flexibility across a number of brain regions
102 predicts individual differences in the speed of acquisition of a simple motor task
103 (Bassett et al., 2011). Network flexibility has also been shown to correlate with
104 working memory and other dimensions of executive function (Braun et al., 2015).
105 But their role in perhaps the most dynamic of cognitive behaviors – trial-by-trial
106 updating of learning based on reinforcement – has not been studied.

107

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

121

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

131 reinforcement learning, we hypothesized (3) that a distinct set of regions would
132 exhibit a relationship between network flexibility and episodic memory.
133
134 To test these hypotheses, we used functional magnetic resonance imaging (fMRI) to
135 measure changes in brain network structure while participants engaged in learning
136 of probabilistic stimulus-outcome associations based on reinforcement (**Figure 1a**).
137 On each trial, participants were presented with a visual cue, then made a choice
138 indicated by a key press, and then received feedback. Associations between cues and
139 outcomes were probabilistic and required updating based on trial-by-trial feedback.
140 We used a task for which behavior has been well described by reinforcement
141 learning models (Foerde and Shohamy, 2011), and which is known from fMRI to
142 involve the striatum (Foerde and Shohamy, 2011) and from patient studies to
143 depend on it (Foerde et al., 2012). To test the role of network dynamics in episodic
144 memory, the task also included trial-unique images presented during feedback. Each
145 of these images coincided with reinforcement, but they were incidental to the
146 learning task (**Figure 1b**). Thus, this task allowed us to test hypotheses about how
147 dynamic changes in network coupling relate to reinforcement learning and to
148 episodic memory.
149

150 Results

151 *Reinforcement Learning Performance.* Participants learned the optimal response for
152 each cue. Percentage of optimal responses increased continuously from 68% in the
153 first block to 76% in the final block, on average. Using a mixed-effects logistic model
154 predicting optimal response with learning block, we observed a significant effect of
155 block on learning performance, as measured by the proportion of optimal responses
156 during each block (**Figure 1c**; $\beta = 0.28$, Standard Error (S.E.) = 0.11, $p = 0.01$ (Wald
157 approximation, (Bates et al., 2015)).
158

159 We also fit standard reinforcement learning models (Daw, 2011; Sutton and Barto,
160 1998) to participants' trial-by-trial choice behavior (**Experimental Procedures**).
161 As expected, the models fit learning behavior significantly better than chance
162 (**Experimental Procedures**, $t_{21}=23.04$, $p<0.0001$).
163

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

177
178 *Brain Network Flexibility.* To probe the role of dynamic brain circuits, we first sought
179 to characterize spatial and temporal properties of brain network dynamics during
180 the task, separately from any information about learning performance. We
181 constructed dynamic functional connectivity networks for each subject in 50 s
182 windows, and used a recently developed multi-slice community detection algorithm
183 (Mucha et al., 2010) to partition each network into dynamic communities: groups of
184 densely connected brain regions that evolve in time (**Experimental Procedures**).
185 In 110 cortical and subcortical ROIs from the Harvard-Oxford atlas, we computed a
186 flexibility statistic for each learning block, which measures the proportion of
187 changes in that region's allegiance to large-scale communities over time (Bassett et
188 al., 2011).

189
190 We began by examining the regional distribution of network flexibility across the
191 brain during learning by computing the average flexibility across the whole learning
192 task (see **Experimental Procedures**). As predicted, sensory and motor regions
193 consistently showed the lowest levels of flexibility (Betzel et al., 2016), while
194 association cortices showed moderate to high levels of flexibility (**Figure 2a**). In
195 addition to this regional distribution, we sought to verify that flexibility was not
196 related to the size of the ROI. We found that ROI size only explained 1.17 % of the
197 variance in average flexibility across individuals, $r = 0.11$, $t_{108} = 1.311$, $p = 0.26$),
198 indicating that this measure is not an artifact of the parcellation we used. We also
199 studied the temporal profile of this measure, by examining changes in flexibility
200 over the course of the task. Whole-brain flexibility (averaged across all ROIs)
201 increased in early learning blocks, before decreasing in later stages of the task
202 (**Figure 2b**).
203

204 *Flexibility in the striatum relates to reinforcement learning.* Previous reports have
205 implicated large-scale network flexibility in motor performance. The temporal
206 profile of network flexibility during this experiment suggests a parallel with
207 learning behavior. To test specifically whether flexibility in the striatum's network
208 coupling is related to learning performance, we fit a mixed effects logistic regression
209 (Bates et al., 2015) using average flexibility in an *a priori* striatum ROI during
210 individual learning blocks to predict performance. Striatal flexibility computed for
211 each block was significantly associated with accuracy on the same block (**Figure 3a**;
212 $\beta = 9.45$, S.E. = 2.75, $p < 0.001$ (Wald approximation (Bates et al., 2015)))
213 indicating that network flexibility is related to behavioral measures of learning. For
214 appropriate posterior inference, we fit a Bayesian extension of this model using Stan
215 (Carpenter et al., 2015) (**Experimental Procedures**), to generate a posterior 95%
216 credible interval (C.I.) of [3.53, 14.99], indicating that flexibility changes of 1% are
217 associated with learning changes of 0.88-3.80% on average (**Supplemental Figure**
218 **1**). To ensure that this relationship reflected a within-subjects effect of flexibility on
219 learning, we included average flexibility across blocks for each subject in the model.
220 This extended model produced similar results ($\beta = 9.79$, S.E. = 2.82, $p = 0.0005$).
221 These results indicate that changes in coupling between the striatum and other
222 brain networks during learning play a significant role in reinforcement learning.

223
224 *Flexibility relates to learning in a distributed set of brain regions.* To explore regions
225 other than the striatum whose dynamic connectivity across networks predicted
226 learning, we performed the above analysis first for the whole brain and then for
227 each of the 110 cortical and subcortical ROIs. These analyses also allowed us to
228 explore the effects of dynamic network coupling in distinct sub-regions of the
229 striatum, as we performed separate tests on the left and right caudate, putamen, and
230 nucleus accumbens.
231
232 Consistent with reports relating whole-brain flexibility to behavior (Bassett et al.,
233 2011; Braun et al., 2015), we found a significant effect of global flexibility averaged
234 across all ROIs on learning (**Supplemental Figure 2**, $\beta = 11.84$, S.E. = 3.91,
235 $p < 0.005$). By testing each ROI separately, we found further evidence for our
236 hypothesis that the striatum flexibly communicates with a number of brain
237 networks in order to integrate stimulus-action-value information. This analysis
238 again revealed a significant effect of flexibility in striatal sub-regions (the right
239 putamen and left caudate) surviving False Discovery Rate (FDR) correction for
240 multiple comparisons. While the global effect might have been driven solely by the
241 striatum, other regions may also play a role. The corrected results, which are
242 presented in **Figure 3b**, indicate that network flexibility in regions of the motor
243 cortex, parietal lobe, and orbital frontal cortex, among others (see **Supplemental**
244 **Table 1** and **Figure 3** for full list and uncorrected map), are also associated with
245 learning from probabilistic feedback.
246
247 *Individual differences in learning rate correlate with whole-brain flexibility.* We next
248 explored the relationship between flexibility and the learning rate α , a parameter in
249 computational models of reinforcement learning, that quantifies the extent to which
250 subjects weigh feedback from each trial when computing the current value of an
251 action given a particular stimulus (Daw, 2011; Sutton and Barto, 1998), as described
252 above. We focused on flexibility averaged across the whole brain and in the *a priori*
253 striatum ROI. Consistent with our hypothesis and with the observation of wide
254 variability in learning rate across subjects, learning rate exhibited a significant
255 negative correlation with whole-brain flexibility (Spearman's correlation coefficient
256 $\rho = -0.60$, $p < 0.005$, **Figure 4**); that is, participants with a lower learning rate
257 (indicating more integration of information across multiple trials) had more
258 flexibility. There was no significant relationship between learning rate and the
259 proportion optimal responses ($\rho = -0.28$, $p = 0.21$), suggesting that this correlation
260 reflected learning per se, and not general performance. Moreover, the relationship
261 between learning rate and flexibility remained significant while controlling for task
262 performance (partial $\rho = -0.62$, $p < 0.005$). We did not detect a relationship between
263 flexibility in the striatal ROI and learning rate (**Supplemental Figure 4**, $\rho = 0.004$,
264 $p > 0.95$).
265
266 Together these results indicate that reinforcement learning involves dynamic
267 changes in network structure, especially in the coupling of distributed regions that
268 may each be involved in different aspects of learning. They also suggest that

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

271
272 *Flexibility in medial cortical regions is associated with episodic memory.* The task also
273 included trial-unique objects presented simultaneously with reinforcement,
274 allowing us to measure subjects' episodic memory, a process thought to rely on
275 distinct cognitive and neural mechanisms to feedback-based incremental learning.
276 We tested whether network flexibility was associated with episodic memory for
277 these trial-unique images, as assessed in a later surprise memory test. Having a
278 measure of episodic memory for the same trials in the same participants allowed us
279 to determine whether the same striatal network dynamics are correlated with any
280 form of learning, or whether these two forms of learning, occurring at the same
281 time, are related to distinct network dynamics.

282 Behaviorally, participants' memory was better than chance (d -prime = 0.93, $t_{21} =$
283 7.27, $p < 0.0001$). Memory performance ("hits") varied across learning blocks,
284 allowing us to assess within-subject associations between network flexibility and
285 behavior (**Figure 5a**). Memory performance was not correlated with incremental
286 learning performance (mixed effects logistic regression, $\beta = 0.41$, Standard Error
287 (S.E.) = 0.51, $p = 0.42$ (Wald approximation)). We tested the effect of flexibility on
288 memory performance (proportion correct) in each of the 110 ROIs. A whole-brain
289 FDR-corrected analysis revealed one region where flexibility was associated with
290 episodic memory, the left paracingulate gyrus. An exploratory uncorrected analysis
291 revealed regions in the medial prefrontal and medial temporal (parahippocampal)
292 cortices where flexibility was associated with episodic memory ($p < 0.05$
293 uncorrected, **Figure 5b**). None of the sub-regions from our *a priori* striatum ROI
294 passed even this low threshold for an effect of flexibility on episodic memory.

295

296 Discussion

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

306

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

315 An interesting hypothesis generated by the current framework is that network
316 flexibility will play a larger role in learning behavior the more that the learning in
317 question relies on widespread information integration, and that this flexibility will
318 be selective to the regions known to participate in this type of learning. Thus,
319 instrumental conditioning paradigms involving complex audio-visual stimuli (Kehoe
320 and Gormezano, 1980), for example, would be expected to associate more strongly
321 with striatal flexibility than the task presented here. In addition, learning that relies
322 on inferences or integration (Wimmer et al., 2012; Wimmer and Shohamy, 2012) or
323 on the encoding of associations across space or time (Eichenbaum, 2000; Staresina
324 and Davachi, 2009), both processes thought to rely on medial temporal and
325 prefrontal processes, should be associated with network flexibility in these regions.
326 This framework provides testable hypotheses to be explored in future work.
327

328 Multiple disorders have been linked to both learning abnormalities (Fineberg et al.,
329 2010; Foerde et al., 2015; Kheirbek et al., 2012; Waltz and Gold, 2007) and changes
330 in diffuse connectivity patterns (Greicius, 2008), with increasing evidence that these
331 changes involve aberrations in connectivity dynamics (Damaraju et al., 2014; Ma et
332 al., 2014). The results presented in this report raise the possibility of linking these
333 separate strands of research to better understand mental diseases with learning
334 components, such as schizophrenia (Waltz and Gold, 2007), anxiety (Kheirbek et al.,
335 2012), and eating disorders (Foerde et al., 2015), among others. In this framework,
336 abnormalities in the integration and routing of information, expressed as
337 differences in dynamic patterns of large-scale functional connectivity (Braun et al.,
338 2016), could lead to the deficiencies in learning hypothesized to represent cognitive
339 endophenotypes for mental disorders (Robbins et al., 2012).
340

341 There are a number of limitations to this report. First, given the static feedback
342 probabilities, the relationship between network flexibility and learning performance
343 could be confounded by the amount time on task for each subject. This seems
344 unlikely to fully explain this relationship, given that the relationship between
345 episodic memory and network flexibility did not increase linearly over time.
346 Nonetheless, future studies incorporating reversal periods to disassociate overall
347 time on task from learning performance will further address this issue. Another
348 limitation is the hard partition for network assignments provided by current multi-
349 slice community detection methods utilized in this study, which necessarily
350 underemphasize uncertainty about community assignment. There have been very
351 recent attempts to formalize probabilistic models of dynamic community structure
352 (Durante et al., 2016; Palla et al., 2016), but most work examining dynamic
353 communities in the brain have used deterministic community assignment (Bassett
354 et al., 2011; Bassett et al., 2015; Shine et al., 2016). More work needs to be done to
355 develop and validate these probabilistic models and apply them to neuroscience
356 data. Relatedly, while the spatial resolution of fMRI makes it an appealing method to
357 begin examining network dynamics, further characterization will entail the use of
358 methods with higher resolution such as ECoG (Khambhati et al., 2016) and MEG
359 (Siebenhühner et al., 2013).
360

361
362 To summarize, we report a novel link between computational models of
363 reinforcement learning, which have often relied on the activity of individual regions
364 to provide an underlying brain mechanism for learning from repeated experiences,
365 with a dynamic measure of network connectivity. Studies of the role of large-scale
366 brain networks in behavior have in contrast often eschewed discussion of the
367 mechanistic roles of single regions, in part because of the challenge of linking
368 widespread connectivity patterns to specific temporal epochs (Medaglia et al.,
369 2015). However, recent advances in network theory are beginning to make the role
370 of dynamic communication between individual regions and broader networks in a
371 host of behaviors a tractable area of research (Bassett et al., 2011; Bassett et al.,
372 2015; Braun et al., 2015; Kopell et al., 2014). Incremental learning based on
373 reinforcement, which necessarily relies on large-scale information integration to
374 inform future behavior, represents a cognitive domain in which such dynamic local-
375 widespread communication provides a rich framework for understanding brain-
376 behavior relationships (Büchel et al., 1999).

377
378 **Materials and Methods**

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

389
390 *Task and Behavioral Analysis.* The probabilistic learning task administered to
391 subjects undergoing an fMRI session has been previously described (Davidow et al.,
392 2016; Foerde et al., 2012; Foerde et al., 2013; Foerde and Shohamy, 2011). Before
393 scanning, participants completed a practice round of 8 trials to become familiar with
394 the task. During the imaging session, individuals underwent an instrumental
395 conditioning procedure, in which they learned to associate 4 cues with two possible
396 outcomes. The cues were images of butterflies; the outcomes were images of
397 flowers. On each trial, participants were presented with an image of one of the four
398 butterflies along with two flowers, and asked to indicate which flower the butterfly
399 was likely to feed from, using a left or right button press. They were then given
400 feedback consisting of the words 'Correct' or 'Incorrect'. Presentation of the
401 feedback also included an image of an object unique to each trial, shown in random
402 order for the purpose of subsequent memory testing. For each butterfly image, one
403 flower represented the 'optimal' choice, with a 0.8 probability of being correct,
404 while the alternative flower had a 0.2 probability of being followed by correct
405 feedback. Subjects performed four blocks of this probabilistic learning phase, each
406 consisting of 30 trials. Feedback was presented for 2 seconds, and was followed by a

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

410
411 For each trial in the learning phase, both the feedback received as well as whether
412 or not subjects made the optimal choice were recorded, and percent correct for each
413 block was computed as the percent of trials on which subjects made the optimal
414 choice, regardless of feedback. These variables enable a characterization of learning
415 as the proportion of optimal choices in each block, as well as that in the test phase.
416 Using this information, we fit reinforcement learning models to subjects' decisions
417 (Daw, 2011; Sutton and Barto, 1998).

418
419 Following the fMRI session (30 minutes), subjects were given a surprise memory
420 test for the trial-unique object images presented during feedback in the learning
421 phase. Subjects were presented with all 120 objects shown during the conditioning
422 phase, along with an equal number of novel objects, and asked to judge the images
423 as "old" or "new". They were also asked to rate their confidence for each decision on
424 a scale of 1-4 (one being most confident; four indicating "guessing"). All responses
425 rated 4 were excluded from our analyses (Foerde and Shohamy, 2011).

426
427 *Reinforcement Learning Model.* To characterize learning on a trial-by-trial basis, we
428 fit standard reinforcement learning models to individuals' choice behavior (Daw,
429 2011; Sutton and Barto, 1998). Briefly, the expected value for a given choice at time
430 t , Q_t , is updated based on the reinforcement outcome r_t via a prediction error δ_t :

431

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

432
433 The reinforcement learning models included two free parameters, α and β . The
434 learning rate α is a parameter between 0 and 1 that measures the extent to which
435 value is updated by feedback from a single trial. Higher α indicates more rapid
436 updating based on few trials and lower α indicates slower updating based on more
437 trials. Another parameter fit to each subject is the inverse temperature parameter β ,
438 which determines the probability of making a particular choice using a softmax
439 function (Daw, 2011; Ishii et al., 2002), so that the probability of choosing choice 1
440 on trial t would be:

441

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

442
443 where $p(c_t = 1)$ refers to the probability of choice one and Q_{1t} is the value for this
444 choice on trial t .

445
446 The two free parameters, α and β , were fit to each participant's behavioral data by
447 maximizing an estimate of the log likelihood using a minimizing optimizer (fmincon
448 in MATLAB) with the negative log likelihood as the error function. To constrain the
449

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

$$\begin{aligned}\beta &\sim \text{Gamma}(1.2, 5) \\ \alpha &\sim \text{Beta}(1.1, 1.1).\end{aligned}$$

454

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

458

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

463

464 *MRI Acquisition.* MRI images were acquired on a 3 T Siemens Tim Trio scanner using
465 a 12-channel head coil. For each block of the learning phase of the conditioning
466 task, we acquired 200 interleaved T2*-weighted echo-planar (EPI) volumes with the
467 following sequence parameters: TR = 2000 ms; TE = 30 ms; flip angle (FA) = 90°;
468 array = 64 x 64; 34 slices; effective voxel resolution = 3x3x4 mm; FOV = 192 mm). A
469 high resolution T1-weighted MPRAGE image was acquired for registration purposes
470 (TR = 2170 ms, TE = 4.33 ms, FA = 7°, array = 256 x 256, 160 slices, voxel resolution
471 = 1 mm³, FOV = 256).

472

473 *fMRI Preprocessing and Dynamic Connectivity Analysis.* Functional images were
474 preprocessed using FSL's FMRI Expert Analysis Tool (FEAT (Smith et al., 2004)).
475 Images from each learning block were high-pass filtered at $f > 0.008$ Hz, spatially
476 smoothed with a 5mm FWHM Gaussian kernel, grand-mean scaled, and motion
477 corrected to their median image using an affine transformation with tri-linear
478 interpolation. The first three images were removed to account for saturation effects.
479 Functional and anatomical images were skull-stripped using FSL's Brain Extraction
480 Tool. Functional images from each block were co-registered to subject's anatomical
481 images and non-linearly transformed to a standard template (T1 Montreal
482 Neurological Institute template, voxel dimensions 2 mm³) using FNIRT (Andersson
483 et al., 2008). Following image registration, time courses were extracted for each
484 block from 110 cortical and subcortical regions of interest (ROIs) segmented from
485 FSL's Harvard-Oxford Atlas. Due to known effects of motion on measures of
486 functional connectivity (Power et al., 2012; Satterthwaite et al., 2012b), time
487 courses were further preprocessed via a nuisance regression. This regression
488 included the six translation and rotation parameters from the motion correction
489 transformation, average CSF, white matter, and whole brain time courses, as well as
490 the first derivatives, squares, and squared derivatives of each of these confound
491 predictors (Satterthwaite et al., 2012a).

492

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

496 0.06-0.12 Hz in order to later assess modularity over short time windows in a
497 manner consistent with previous reports (Bassett et al., 2011; Bassett et al., 2013b):

498
$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)} ,$$

499 where $G_{xy}(f)$ is the cross-spectral density between regions x and y, and $G_{xx}(f)$ and
500 $G_{yy}(f)$ are the autospectral densities of signals x and y, respectively. We thus
501 created subject-specific $110 \times 110 \times 8$ connectivity matrices for 110 regions and 8
502 time windows for each of the 4 learning blocks, containing coherence values ranging
503 between 0 and 1. The frequency range of 0.06-0.12 Hz was chosen to approximate
504 the frequency envelope of the hemodynamic response, allowing us to detect changes
505 as slow as 3 cycles per window with a 2 second TR.

506

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

524

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

540 non-trivial re-arrangements during motor skill acquisition (Bassett et al., 2011;
541 Bassett et al., 2015) and memory processing (Braun et al., 2015). These studies
542 support the utility of module-based analyses in the examination of higher order
543 cognitive processes in functional neuroimaging data.
544
545 To extract modules or communities from a single-network representation, one
546 typically applies a community detection technique such as modularity maximization
547 (Newman, 2004). However, these single-network algorithms do not allow for the
548 linking of communities across time-points, thus hampering statistically robust
549 inference regarding the reconfiguration of communities as the system evolves
550 (Mucha et al., 2010). In contrast, the multilayer approaches introduced above allow
551 for the characterization of multi-layer network modularity, with layers representing
552 time windows within a learning block. In this framework, each network node (or
553 ROI) in the multi-layer network is connected to itself in the preceding and following
554 time windows in order to link networks in time. This organization enables us to
555 solve the famous community-matching problem explicitly within the model (Mucha
556 et al., 2010), and in principle also facilitates the examination of module
557 reconfiguration across multiple temporal resolutions of system dynamics (Bassett et
558 al., 2013a). We thus constructed multilayer networks for each subject for each of the
559 four learning blocks in the experiment, allowing for the partitioning of each network
560 into communities or modules whose identity is robustly tracked across time
561 windows.
562
563 The partitioning of these multilayer networks into temporally linked communities
564 was carried out using a Louvain-like locally greedy algorithm for multilayer
565 modularity optimization (Jutla et al., 2011; Mucha et al., 2010).

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

566 where Q_{ml} is the multilayer modularity index. The adjacency matrix for each layer l
567 consists of components A_{ijl} . The variable γ_l represents the resolution parameter for
568 layer l , while C_{jlr} gives the coupling strength between node j at layers l and r (see
569 below for details of fitting these two parameters). The variables g_{il} and g_{jr}
570 correspond to the community labels for node i at layer l and node j at layer r ,
571 respectively; k_{il} is the connection strength (in this case, coherence) of node i in layer
572 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,
573 the function $\delta(g_{il}, g_{jr})$ refers to the Kronecker delta function, which equals 1 if
574 $g_{il}=g_{jr}$, and 0 otherwise.
575

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

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

592
593 Finally, we note that maximization of the modularity quality function is NP-hard,
594 and the Louvain-like locally greedy algorithm we employ is a computational
595 heuristic with non-deterministic solutions. Due to the well known near-degeneracy
596 of Q_{ml} (Bassett et al., 2013a; Good et al., 2010; Mucha et al., 2010), we repeated the
597 multi-slice community detection algorithm 500 times using the resolution and
598 coupling parameters estimated from the grid search procedure outlined above. This
599 approach ensured an adequate sampling of the null distribution (Bassett et al.,
600 2013a). Each repetition produced a hard partition of nodes into communities as a
601 function of time window: that is, a community allegiance identity for each of the 110
602 brain regions in the multilayer network.

603
604 *Flexibility –A Dynamic Network Statistic.* To characterize the dynamics of these
605 temporal networks and their relation to learning, we computed the *flexibility* of each
606 node, which measures the extent to which a region changed its community
607 allegiance over time (Bassett et al., 2011) . Intuitively and more colloquially,
608 flexibility can be thought of as a measure of a region's tendency to communicate
609 with different networks during learning. Flexibility for a learning block is defined as
610 the number of times a node displays a change in community assignment over time,
611 divided by the number of possible changes (which is equal to the number of time
612 points in that block minus 1). To obtain a stable estimate of flexibility, we averaged
613 flexibility scores for each ROI over the 500 iterations of the multilayer community
614 detection algorithm described above. This measure was computed for each region in
615 each block. In addition, average measures of flexibility were computed across the
616 whole-brain and across all blocks. To assess the spatial distribution of flexibility
617 over the brain, we visualized the average flexibility over all blocks and subjects for
618 each region using surface maps.

619
620 *Relating Flexibility to Reinforcement Learning.* To examine the effect of flexibility on
621 learning from feedback, we estimated a generalized mixed-effects model predicting
622 optimally correct choices with flexibility estimates for each block with a logistic link
623 function, using the Maximum Likelihood (ML) approximation implemented in the
624 *lme4* package (Bates et al., 2015). Subjects' average flexibility in an *a priori* striatum
625 ROI was used to predict the proportion of optimal choices in each learning block.
626 The ROI included bilateral caudate, putamen, and nucleus accumbens regions from
627 the Harvard-Oxford atlas. We included a random effect of subject, allowing for

628 different effects of flexibility on learning for each subject, while constraining these
629 effects with the group average. Average flexibility across sessions was included as a
630 fixed effect in the model in order to ensure that our estimates represented within-
631 subject learning effects. We also estimated this effect for whole-brain flexibility,
632 which has related to several cognitive functions in previous reports (Bassett et al.,
633 2011; Braun et al., 2015).

634
635 To provide appropriate posterior inference about the plausible parameter values
636 indicated by our data, and to account for uncertainty about all parameters, we also
637 fit a fully Bayesian extension of the ML approximation described above for the effect
638 of striatal flexibility on learning performance (**Supplemental Figure 1**). We used
639 the ‘brms’ package for fitting models in the Stan language (Carpenter et al., 2015).
640 These models were similar to the likelihood approximation models, but included a
641 covariance parameter for subject-level slopes and intercepts (which could not be fit
642 by the above approximation), and weakly informative prior distributions to
643 regularize parameter estimation:

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

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

644
645 where β represents the “fixed effects” parameters (slope and intercept), τ
646 represents the “random effects” variance for subject-level estimates sampled from
647 β , and T^+ is a positive half-t distribution (Gelman, 2006). We used an lkj(2) prior for
648 correlations between subject-level intercept and slope estimates (Lewandowski et
649 al., 2009).
650

651
652 To explore other regions exhibiting effects of dynamic connectivity on learning
653 performance, we separately modeled the effect of flexibility in each brain region on
654 reinforcement learning using the ML approximation implemented in the *lme4*
655 package. To control for Type I error, we applied a false discovery rate correction for
656 multiple comparisons across regions (Benjamini and Hochberg, 1995). While
657 regions passing this threshold are reported, we also visualize the results using an
658 exploratory uncorrected threshold of $p < 0.05$ (**Supplemental Figure 3**).
659

660
661 To examine the relationship between flexibility and parameters estimated from
662 standard reinforcement learning models, we tested whether whole-brain and
663 striatal flexibility were correlated with the learning rate α for each subject, using
664 Spearman’s correlation coefficient due to the non-Gaussian distribution of learning
665 rates. If flexibility is related to the integration of sensory, motor, and value
666 information over time, it should be associated with learning rate, which measures
667 the extent to which subjects average feedback from individual trials when updating
668 instrumental value, as opposed to averaging over many experiences. Lower values
669 of the learning rate indicate integration of information across more trials; higher
670 values of learning rate indicate integration of information across fewer trials.
671

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

673 scores for the trial-unique objects presented during feedback. If the effects of
674 striatal flexibility are relatively selective to incremental learning, we expected to
675 find no significant association even at an uncorrected threshold with memory in the
676 regions comprising our striatal ROI. In addition, this provided an exploratory
677 analysis to examine the regions in which network flexibility plays a potential role in
678 episodic memory. Given a host of previous studies on multiple learning systems, we
679 reasoned it might be possible to detect an effect of dynamic network coupling on
680 episodic memory in regions traditionally associated with this form of learning.
681

682

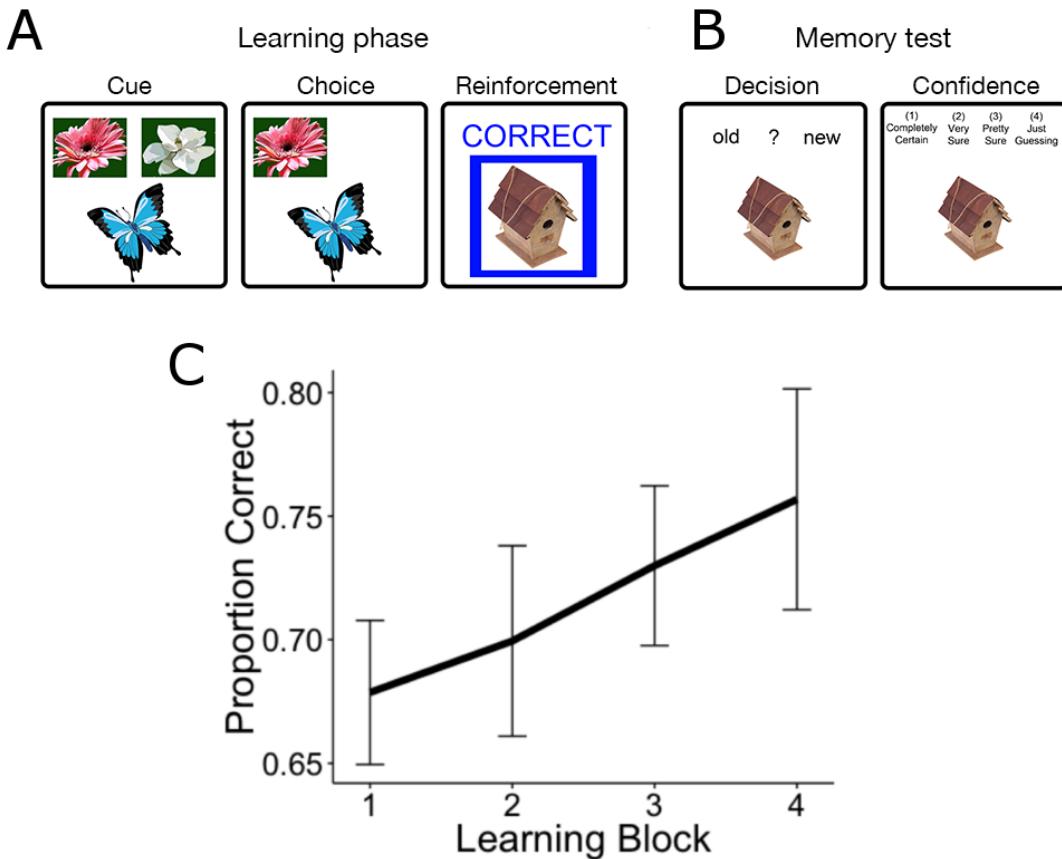
683

684 Acknowledgements: DSB acknowledges support from the John D. and Catherine T.
685 MacArthur Foundation, the Alfred P. Sloan Foundation, the Army Research
686 Laboratory and the Army Research Office through contract numbers W911NF-10-2-
687 0022 and W911NF-14-1-0679, the National Institute of Mental Health (2-R01-DC-
688 009209-11), the National Institute of Child Health and Human Development
689 (1R01HD086888-01), the Office of Naval Research, and the National Science
690 Foundation (CRCNS award BCS-1441502 and CAREER award PHY-1554488). AG
691 acknowledges support from the William T. Grant Foundation and the Center for
692 Translational and Prevention Science (P30 DA027827, Brody-PI) funded by NIDA.
693 D.S. acknowledges support from the National Science Foundation (Career Award
694 #0955494) and the National Institute of Health (NINDS grant # R01NS079784 and
695 CRCNS grant #R01DA038891).

696

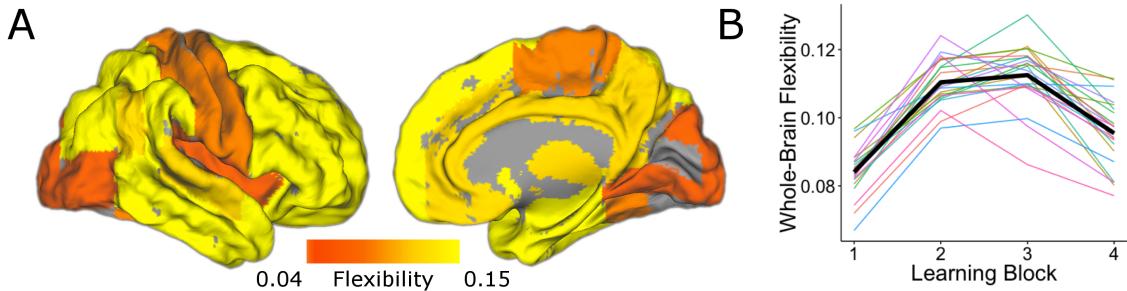
697

698 Figures



699

700 **Figure 1. Task design and learning performance.** Participants performed a
701 modified reinforcement learning task while undergoing fMRI(Foerde and Shohamy,
702 2011). **A.** Participants were instructed to associate each of 4 cues (butterflies) with
703 one of two outcomes (flowers). Feedback was probabilistic, with positive feedback
704 following the choice on 80% of correct trials and on 20% of incorrect trials. **B.** Each
705 feedback event was presented with a unique image. Thirty minutes following the
706 MRI scan, participants were given a surprise episodic memory test, testing
707 recognition and confidence for images seen during the scan, intermixed with novel
708 images. **C.** Average performance on the learning task improved linearly, suggesting
709 continuous learning across all trials.
710



711
712 **Figure 2. Spatial and temporal characteristics of network flexibility.** Network
713 flexibility exhibited distinct spatial and temporal patterns. **A.** Flexibility averaged
714 across all learning blocks was highest across regions of association cortex and
715 lowest in sensory and motor regions. **B.** Flexibility averaged across the whole brain
716 consistently increased in early learning blocks. Color lines represent whole-brain
717 flexibility for single subjects.
718

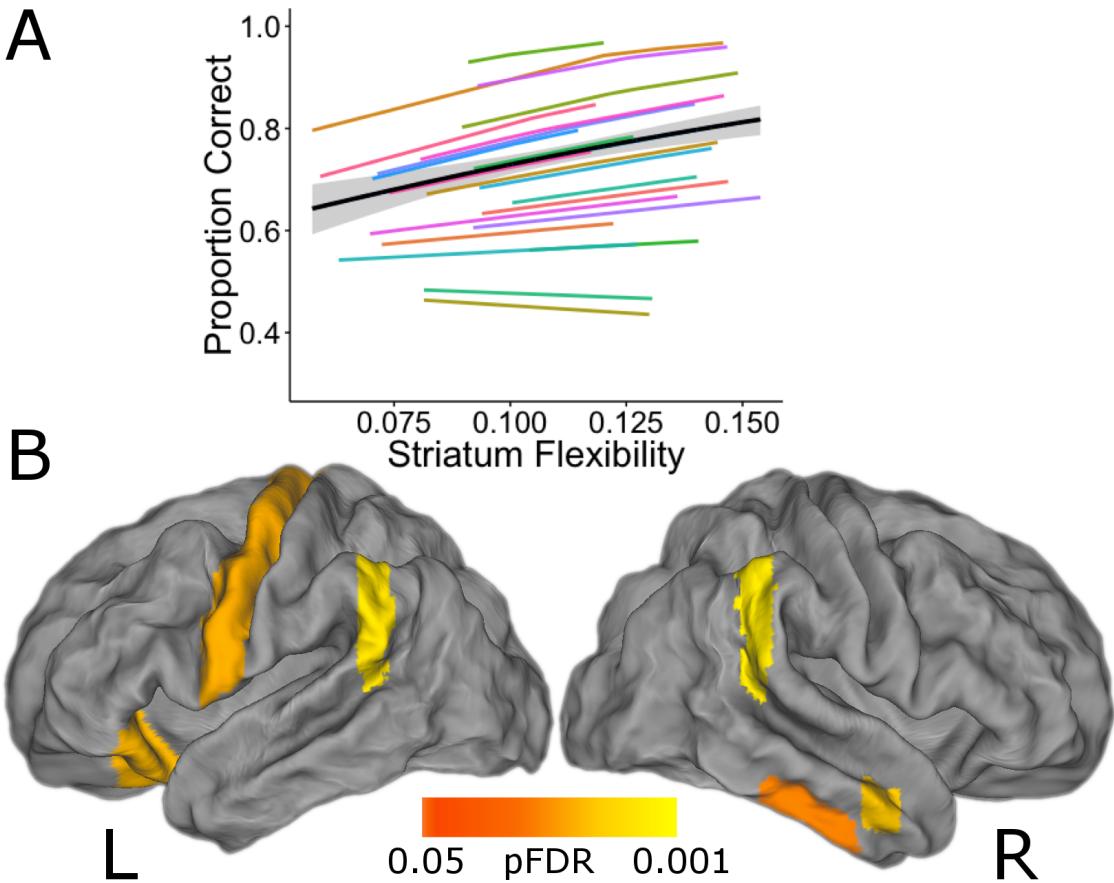
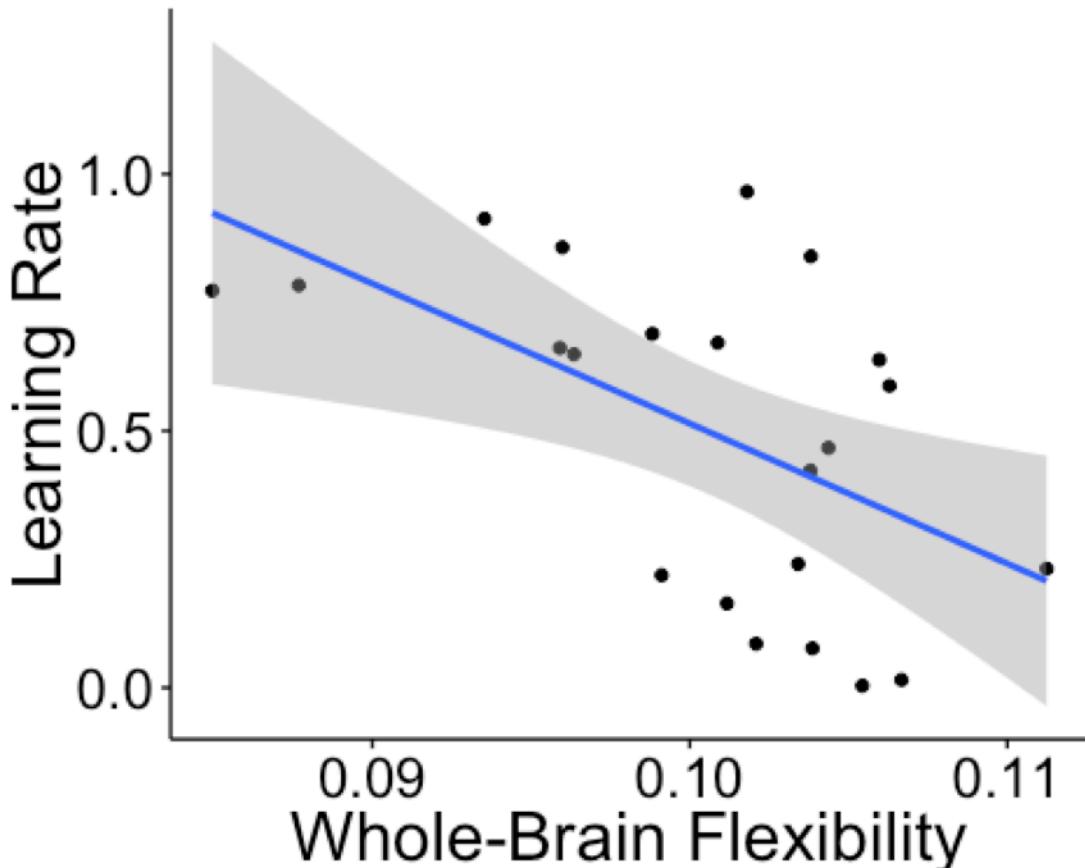
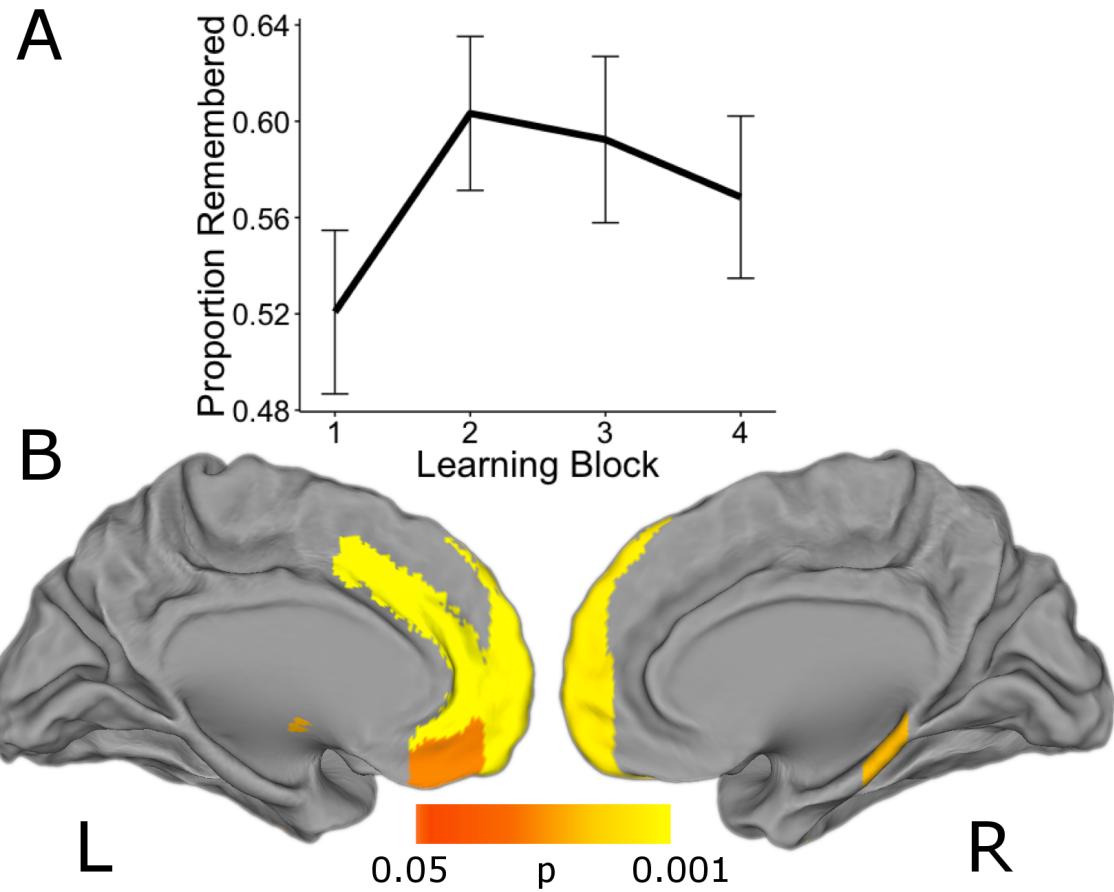


Figure 3. Flexibility in the striatum and beyond relates to learning performance. Network flexibility in a number of brain regions is related to learning performance within subjects. **A.** Mixed-effects model fit for the association between network flexibility in an *a priori* striatum ROI and learning 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.



732
733 **Figure 4. Whole-brain flexibility is negatively correlated with single-trial**
734 **learning rate.** Standard reinforcement learning models were fit to each
735 participant's choice behavior. The learning rate parameter from these models, α ,
736 indexes subjects' tendency to weight single trials (high α) as opposed to averaging
737 over many experiences (low α), and was negatively correlated with whole-brain
738 flexibility. These results indicate that whole-brain flexibility is greater when
739 individuals are integrating information across many trials. Grey band represents
740 95% confidence interval.



741
742 **Figure 5. Network flexibility in medial prefrontal cortex relates to episodic**
743 **memory.** An exploratory analysis showed effects of network flexibility on episodic
744 memory performance in medial prefrontal and temporal lobes. **A.** Average memory
745 (% correct) across trials. Participants' recollection accuracy varied across blocks.
746 Line represents group average and bars represent standard errors. **B.** A number of
747 medial prefrontal regions as well as the right parahippocampal gyrus passed an
748 exploratory uncorrected threshold of $p < 0.05$ for the effect of flexibility on
749 subsequent episodic memory. The effect in the left paracingulate gyrus survived
750 FDR correction.

752 **References**

- 753 Alexander, G.E., DeLong, M.R., and Strick, P.L. (1986). Parallel organization of
754 functionally segregated circuits linking basal ganglia and cortex. Annual review of
755 neuroscience 9, 357-381.
- 756 Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., and Calhoun, V.D. (2012).
757 Tracking whole-brain connectivity dynamics in the resting state. Cerebral cortex,
758 bhs352.
- 759 Andersson, J., Smith, S., and Jenkinson, M. (2008). Fnirt-fmrib's non-linear image
760 registration tool. Human Brain Mapping 2008.
- 761 Bassett, D.S., Porter, M.A., Wymbs, N.F., Grafton, S.T., and Carlson, J.M. (2013a).
762 Robust detection of dynamic community structure in networks. Chaos 23, 013142.
- 763 Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., and Grafton, S.T.
764 (2011). Dynamic reconfiguration of human brain networks during learning.
765 Proceedings of the National Academy of Sciences 108, 7641-7646.
- 766 Bassett, D.S., Wymbs, N.F., Rombach, M.P., Porter, M.A., Mucha, P.J., and Grafton, S.T.
767 (2013b). Task-based core-periphery organization of human brain dynamics. PLoS
768 computational biology 9, e1003171.
- 769 Bassett, D.S., Yang, M., Wymbs, N.F., and Grafton, S.T. (2015). Learning-induced
770 autonomy of sensorimotor systems. Nature Neuroscience.
- 771 Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects
772 Models Usinglme4. Journal of Statistical Software 67.
- 773 Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a
774 practical and powerful approach to multiple testing. Journal of the Royal Statistical
775 Society Series B (Methodological), 289-300.
- 776 Betzel, R.F., Satterthwaite, T.D., Gold, J.I., and Bassett, D.S. (2016). A positive mood, a
777 flexible brain. arXiv preprint arXiv:160107881.
- 778 Bogacz, R., and Gurney, K. (2007). The basal ganglia and cortex implement optimal
779 decision making between alternative actions. Neural computation 19, 442-477.
- 780 Braun, U., Schäfer, A., Bassett, D.S., Rausch, F., Schweiger, J.I., Bilek, E., Erk, S.,
781 Romanczuk-Seiferth, N., Grimm, O., and Geiger, L.S. (2016). Dynamic brain network
782 reconfiguration as a potential schizophrenia genetic risk mechanism modulated by
783 NMDA receptor function. Proceedings of the National Academy of Sciences 113,
784 12568-12573.
- 785 Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L.,
786 Schweiger, J.I., Grimm, O., Heinz, A., and Tost, H. (2015). Dynamic reconfiguration of
787 frontal brain networks during executive cognition in humans. Proceedings of the
788 National Academy of Sciences 112, 11678-11683.
- 789 Bressler, S.L., and Menon, V. (2010). Large-scale brain networks in cognition:
790 emerging methods and principles. Trends in cognitive sciences 14, 277-290.
- 791 Büchel, C., Coull, J., and Friston, K. (1999). The predictive value of changes in
792 effective connectivity for human learning. Science 283, 1538-1541.
- 793 Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical
794 analysis of structural and functional systems. Nature Reviews Neuroscience 10, 186-
795 198.
- 796 Bullmore, E.T., and Bassett, D.S. (2011). Brain graphs: graphical models of the
797 human brain connectome. Annual review of clinical psychology 7, 113-140.

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

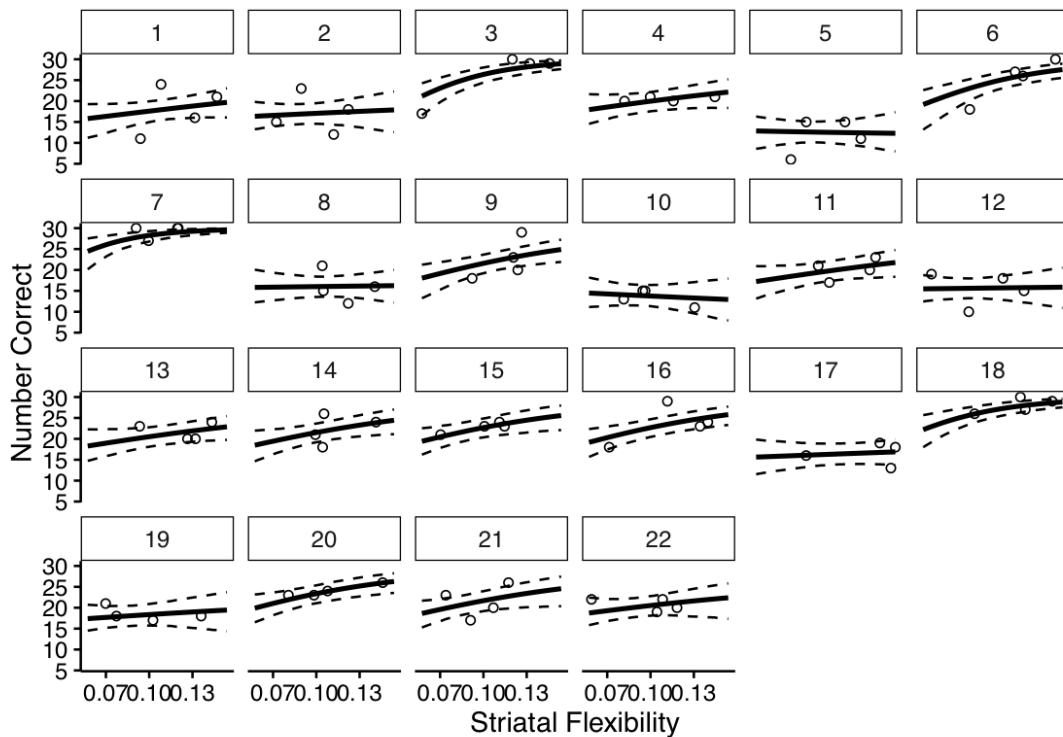
- 844 Frank, M.J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T.V., Cavanagh, J.F., and Badre, D.
845 (2015). fMRI and EEG predictors of dynamic decision parameters during human
846 reinforcement learning. *The Journal of Neuroscience* 35, 485-494.
- 847 Frank, M.J., Seeberger, L.C., and O'reilly, R.C. (2004). By carrot or by stick: cognitive
848 reinforcement learning in parkinsonism. *Science* 306, 1940-1943.
- 849 Gelman, A. (2006). Prior distributions for variance parameters in hierarchical
850 models (comment on article by Browne and Draper). *Bayesian analysis* 1, 515-534.
- 851 Gerraty, R.T., Davidow, J.Y., Wimmer, G.E., Kahn, I., and Shohamy, D. (2014). Transfer
852 of learning relates to intrinsic connectivity between hippocampus, ventromedial
853 prefrontal cortex, and large-scale networks. *The Journal of Neuroscience* 34, 11297-
854 11303.
- 855 Good, B.H., de Montjoye, Y.-A., and Clauset, A. (2010). Performance of modularity
856 maximization in practical contexts. *Physical Review E* 81, 046106.
- 857 Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric
858 disorders. *Current opinion in neurology* 21, 424-430.
- 859 Haber, S.N. (2003). The primate basal ganglia: parallel and integrative networks.
860 *Journal of chemical neuroanatomy* 26, 317-330.
- 861 Haber, S.N., Kim, K.-S., Mailly, P., and Calzavara, R. (2006). Reward-related cortical
862 inputs define a large striatal region in primates that interface with associative
863 cortical connections, providing a substrate for incentive-based learning. *The Journal*
864 of *Neuroscience* 26, 8368-8376.
- 865 Haber, S.N., and Knutson, B. (2010). The reward circuit: linking primate anatomy
866 and human imaging. *Neuropsychopharmacology* 35, 4-26.
- 867 Harmelech, T., Preminger, S., Wertman, E., and Malach, R. (2013). The day-after
868 effect: long term, Hebbian-like restructuring of resting-state fMRI patterns induced
869 by a single epoch of cortical activation. *The Journal of Neuroscience* 33, 9488-9497.
- 870 Hikosaka, O., Kim, H.F., Yasuda, M., and Yamamoto, S. (2014). Basal ganglia circuits
871 for reward value-guided behavior. *Annual review of neuroscience* 37, 289.
- 872 Holme, P., and Saramäki, J. (2011). Temporal Networks. *Physics Reports* 519, 97-
873 125.
- 874 Ishii, S., Yoshida, W., and Yoshimoto, J. (2002). Control of exploitation-exploration
875 meta-parameter in reinforcement learning. *Neural networks* 15, 665-687.
- 876 Jutla, I.S., Jeub, L.G., and Mucha, P.J. (2011). A generalized Louvain method for
877 community detection implemented in MATLAB. URL <http://netwiki.amath.unc.edu/GenLouvain>.
- 878 Kehoe, E.J., and Gormezano, I. (1980). Configuration and combination laws in
879 conditioning with compound stimuli. *Psychological Bulletin* 87, 351.
- 880 Kemp, J.M., and Powell, T. (1971). The connexions of the striatum and globus
881 pallidus: synthesis and speculation. *Philosophical Transactions of the Royal Society*
882 of London B: Biological Sciences
- 883 262, 441-457.
- 884 Khambhati, A.N., Davis, K.A., Lucas, T.H., Litt, B., and Bassett, D.S. (2016). Virtual
885 cortical resection reveals push-pull network control preceding seizure evolution.
886 *Neuron* 91, 1170-1182.
- 887 Kheirbek, M.A., Klemenhagen, K.C., Sahay, A., and Hen, R. (2012). Neurogenesis and
888 generalization: a new approach to stratify and treat anxiety disorders. *Nature*
889 *neuroscience* 15, 1613-1620.

- 890 Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., and Porter, M.A.
891 (2014). Multilayer networks. *Journal of Complex Networks* 2, 203 - 271.
892 Kopell, N.J., Gritton, H.J., Whittington, M.A., and Kramer, M.A. (2014). Beyond the
893 connectome: the dynome. *Neuron* 83, 1319-1328.
894 Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation
895 matrices based on vines and extended onion method. *Journal of multivariate*
896 *analysis* 100, 1989-2001.
897 Liu, X., and Duyn, J.H. (2013). Time-varying functional network information
898 extracted from brief instances of spontaneous brain activity. *Proceedings of the*
899 *National Academy of Sciences* 110, 4392-4397.
900 Ma, S., Calhoun, V.D., Phlypo, R., and Adali, T. (2014). Dynamic changes of spatial
901 functional network connectivity in healthy individuals and schizophrenia patients
902 using independent vector analysis. *NeuroImage* 90, 196-206.
903 Medaglia, J.D., Lynall, M.-E., and Bassett, D.S. (2015). Cognitive Network
904 Neuroscience. *Journal of cognitive neuroscience*.
905 Meunier, D., Lambiotte, R., and Bullmore, E.T. (2010). Modular and hierarchically
906 modular organization of brain networks. *Frontiers in neuroscience* 4, 200.
907 Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., and Onnela, J.-P. (2010).
908 Community structure in time-dependent, multiscale, and multiplex networks.
909 *Science* 328, 876-878.
910 Newman, M.E. (2004). Fast algorithm for detecting community structure in
911 networks. *Physical review E* 69, 066133.
912 O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal
913 difference models and reward-related learning in the human brain. *Neuron* 38, 329-
914 337.
915 Palla, K., Caron, F., and Teh, Y.W. (2016). Bayesian nonparametrics for Sparse
916 Dynamic Networks. *arXiv preprint arXiv:160701624*.
917 Porter, M.A., Onnela, J.-P., and Mucha, P.J. (2009). Communities in networks. *Notices*
918 *of the AMS* 56, 1082-1097.
919 Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2012).
920 Spurious but systematic correlations in functional connectivity MRI networks arise
921 from subject motion. *Neuroimage* 59, 2142-2154.
922 Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C.,
923 Laumann, T.O., Miezin, F.M., and Schlaggar, B.L. (2011). Functional network
924 organization of the human brain. *Neuron* 72, 665-678.
925 Robbins, T.W., Gillan, C.M., Smith, D.G., de Wit, S., and Ersche, K.D. (2012).
926 Neurocognitive endophenotypes of impulsivity and compulsivity: towards
927 dimensional psychiatry. *Trends in cognitive sciences* 16, 81-91.
928 Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E.,
929 Eickhoff, S.B., Hakonarson, H., Gur, R.C., and Gur, R.E. (2012a). An Improved
930 Framework for Confound Regression and Filtering for Control of Motion Artifact in
931 the Preprocessing of Resting-State Functional Connectivity Data. *NeuroImage*.
932 Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H.,
933 Gur, R.C., and Gur, R.E. (2012b). Impact of in-scanner head motion on multiple
934 measures of functional connectivity: relevance for studies of neurodevelopment in
935 youth. *Neuroimage*.

- 936 Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction
937 and reward. *Science* 275, 1593-1599.
- 938 Shine, J.M., Bissett, P.G., Bell, P.T., Koyejo, O., Balsters, J.H., Gorgolewski, K.J., Moodie,
939 C.A., and Poldrack, R.A. (2016). The Dynamics of Functional Brain Networks:
940 Integrated Network States during Cognitive Task Performance. *Neuron* 92, 544-554.
- 941 Shohamy, D., Myers, C., Grossman, S., Sage, J., Gluck, M., and Poldrack, R. (2004).
942 Cortico - striatal contributions to feedback - based learning: converging data from
943 neuroimaging and neuropsychology. *Brain* 127, 851-859.
- 944 Siebenhühner, F., Weiss, S.A., Coppola, R., Weinberger, D.R., and Bassett, D.S. (2013).
945 Intra-and inter-frequency brain network structure in health and schizophrenia. *PloS
946 one* 8, e72351.
- 947 Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-
948 Berg, H., Bannister, P.R., De Luca, M., Dronjak, I., and Flitney, D.E. (2004). Advances
949 in functional and structural MR image analysis and implementation as FSL.
950 *Neuroimage* 23, S208-S219.
- 951 Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann,
952 C.F., Jenkinson, M., Andersson, J., and Glasser, M.F. (2012). Temporally-independent
953 functional modes of spontaneous brain activity. *Proceedings of the National
954 Academy of Sciences* 109, 3131-3136.
- 955 Staresina, B.P., and Davachi, L. (2009). Mind the gap: binding experiences across
956 space and time in the human hippocampus. *Neuron* 63, 267-276.
- 957 Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: An introduction (MIT
958 press).
- 959 Tambini, A., Ketz, N., and Davachi, L. (2010). Enhanced brain correlations during
960 rest are related to memory for recent experiences. *Neuron* 65, 280-290.
- 961 Wagenmakers, E.-J., and Farrell, S. (2004). AIC model selection using Akaike weights.
962 *Psychonomic bulletin & review* 11, 192-196.
- 963 Waltz, J.A., and Gold, J.M. (2007). Probabilistic reversal learning impairments in
964 schizophrenia: further evidence of orbitofrontal dysfunction. *Schizophrenia
965 research* 93, 296-303.
- 966 Wiecki, T.V., and Frank, M.J. (2013). A computational model of inhibitory control in
967 frontal cortex and basal ganglia. *Psychological Review* 120, 329.
- 968 Wimmer, G.E., Daw, N.D., and Shohamy, D. (2012). Generalization of value in
969 reinforcement learning by humans. *European Journal of Neuroscience* 35, 1092-
970 1104.
- 971 Wimmer, G.E., and Shohamy, D. (2012). Preference by Association: How Memory
972 Mechanisms in the Hippocampus Bias Decisions. *Science* 338, 270-273.
- 973 Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M.,
974 Roffman, J.L., Smoller, J.W., Zöllei, L., and Polimeni, J.R. (2011). The organization of
975 the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of
976 neurophysiology* 106, 1125-1165.
- 977
- 978
- 979
- 980

981 **Supplemental Figures**

982



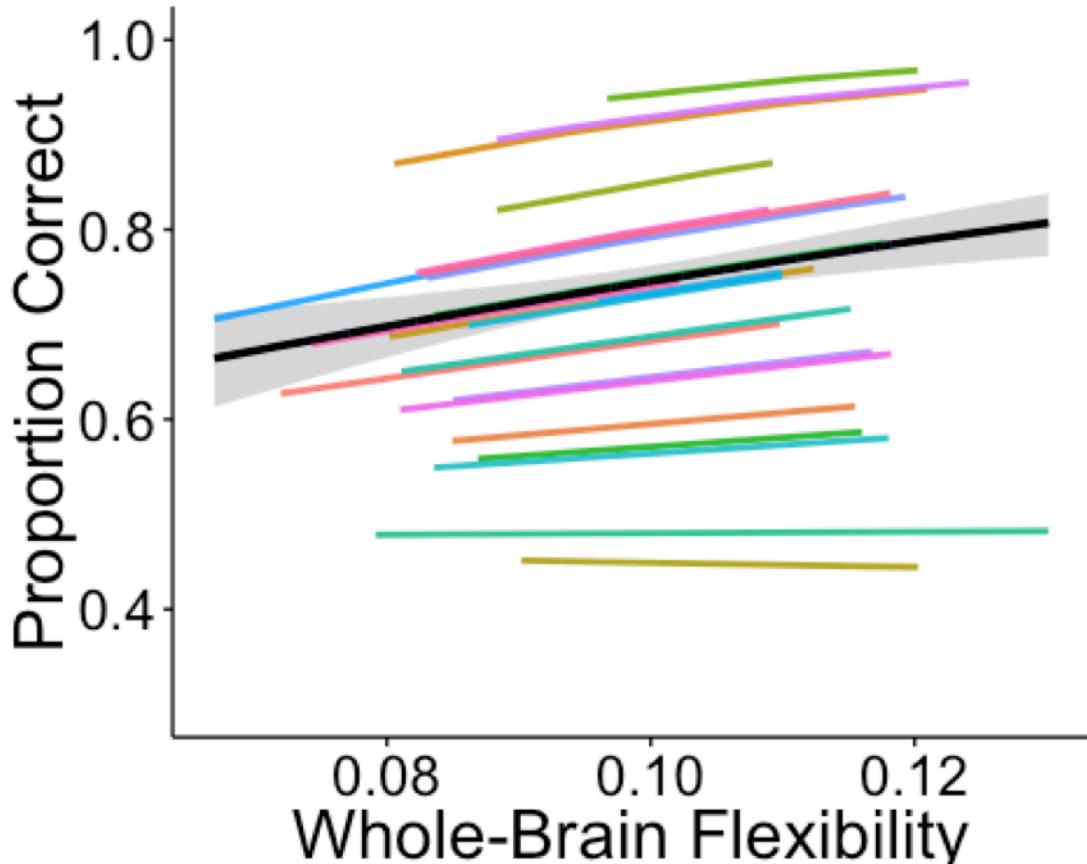
983

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

990

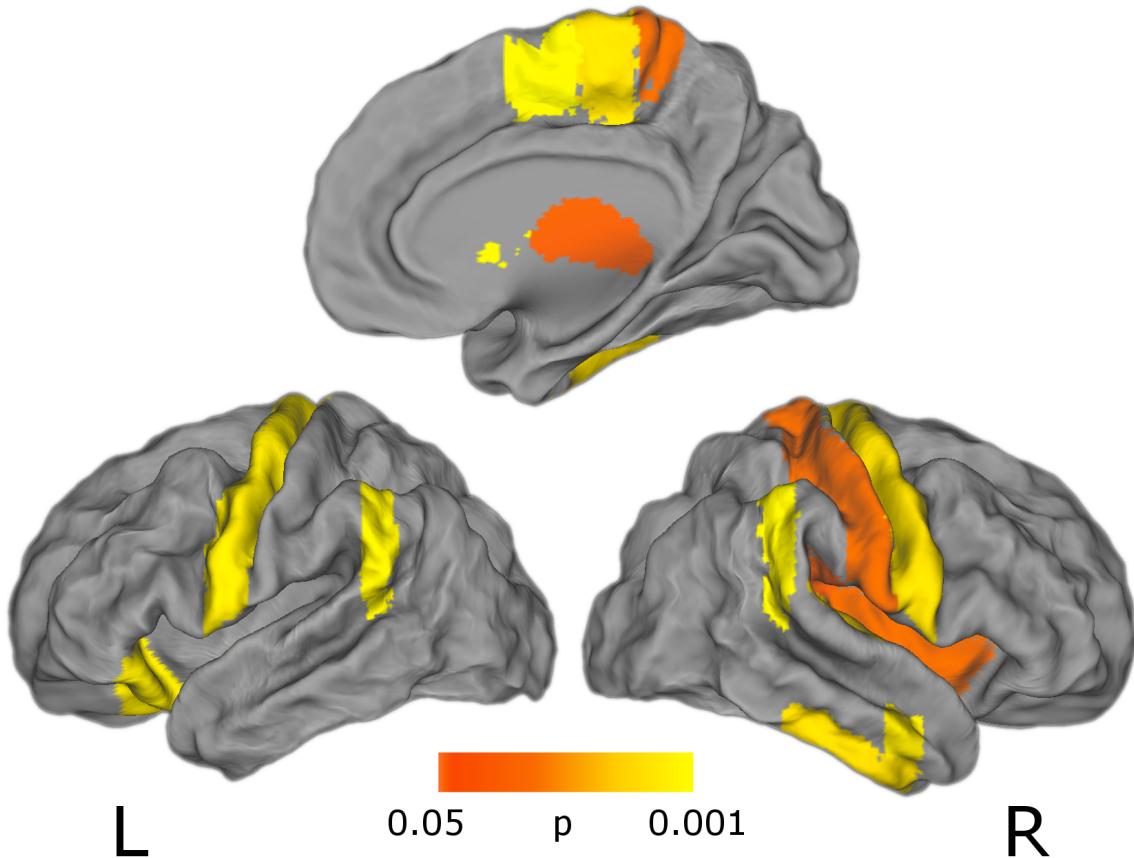
991

992



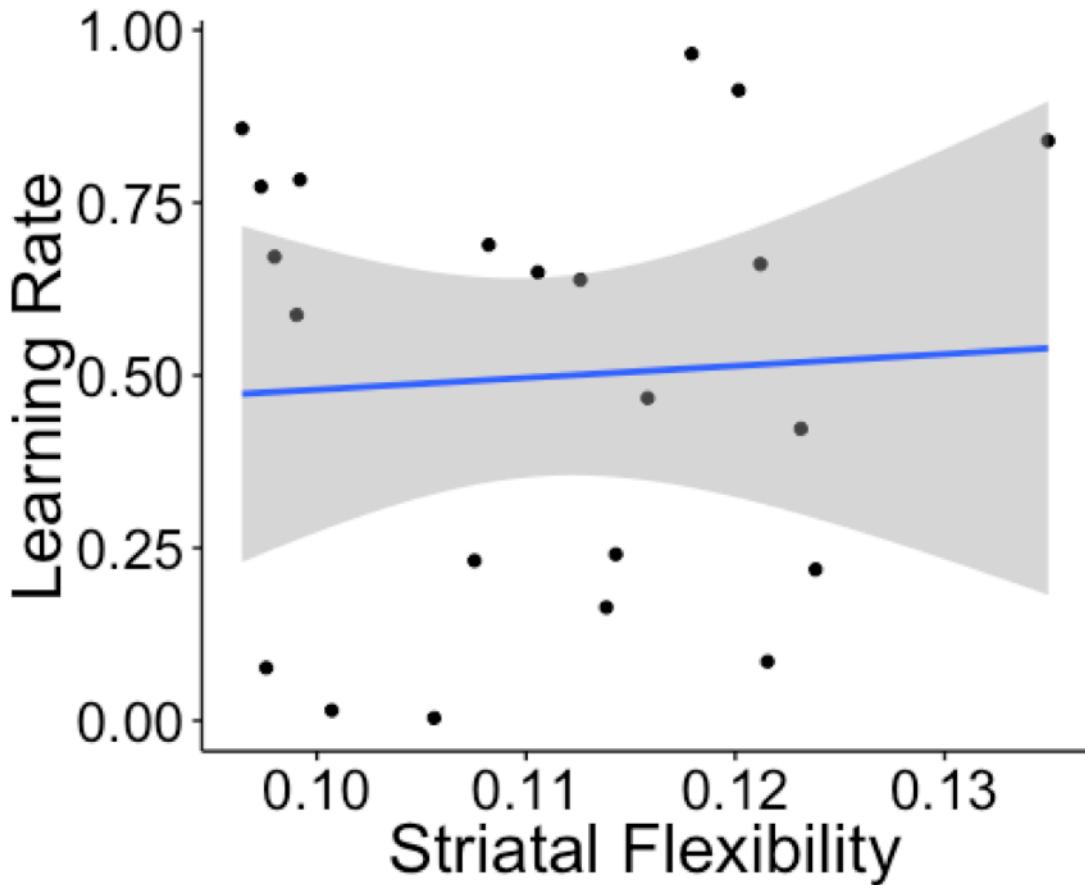
993
994 **Supplemental Figure 2. Whole-brain network flexibility is related to learning**
995 **performance.** Given the distributed effect of flexibility on reinforcement learning,
996 and following other work on motor skills and working memory, we fit a mixed
997 effects model predicting learning performance with a global measure of flexibility
998 averaged across the whole brain. The black line represents the fixed effect estimate
999 and the gray band represents the 95% confidence interval for this estimate, while
1000 color lines represent subject-level random effects estimates.
1001
1002
1003

1004



1005

1006 **Supplemental Figure 3.** Exploratory uncorrected ($p < 0.05$) results for whole-brain
1007 effect of network flexibility on reinforcement learning performance.
1008



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.

1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031

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