Reinforcement Learning and Dynamic Network Flexibility in Adolescence

Chelsea M. Harmon, Raphael T. Gerraty, Juliet Y. Davidow, Karin Foerde, Adriana Galvan, Danielle S. Bassett, Daphna Shohamy

May 21, 2018

AsPredicted Preregistration (OSF Project: Dynamic Flexibility in Adolescence)

What's the main question being asked or hypothesis being tested in this study? (optional)

1. Does greater striatal-cortical dynamic flexibility explain better reinforcement learning performance in adolescents?

Complex learning behaviors involve the integrated action of distributed brain circuits. One example is reinforcement learning that requires the orchestration of several key brain circuits including those that are pertinent for reward valuation, decision making, and memory to integrate visual, motor, and cognitive information. Importantly, in this type of learning paradigm, choices must be updated over the course of the task as reinforcement information accumulates. Moreover, previous research highlights that time-varying changes in patterns of functional connectivity can reflect the dynamics of learning task demands. By formally encoding such changes in network models and applying tools from graph theory, prior work has demonstrated that dynamic network flexibility -DNF- tracks the course of motor learning (Bassett et al. 2011) and is a proposed mechanism for information integration necessary for effective reinforcement learning (Gerraty et al. 2018) in adults. Further, it has recently been proposed that dynamic changes in communication between the striatum and distributed brain networks provide a mechanism for information integration during reinforcement learning in adults (Gerraty et al. 2018). Specifically, flexibility in striatal-cortical circuits relates to learning performance, and whole brain dynamic network flexibility predicts individual differences in parameters from reinforcement learning models.

Adolescence is a unique period marked by sensitivity to reward (Cohen et al. 2010) and poor impulse control resulting from differential neural patterns across the brain (see (Casey, Jones, and Hare 2008) for review). As such, in an adolescent sample, 13-17 year olds have demonstrated better learning performance on a reinforcement learning task compared to an adult sample (Davidow et al. 2016). The ventral striatum has been consistently implicated as sensitive to unexpected positive feedback as reflected by model-based reward prediction error signals (Pagnoni et al. 2002; Cohen et al. 2010). However, how neural dynamic network flexibility may support the integration of information involved in learning and updating over time during adolescence has remained elusive.

Previous literature has largely focused on resting state functional connectivity networks across development (Gu et al. 2015; Fair, et al . 2007); however, the role of dynamic network flexibility -DNF- during an activity, such as active learning and updating information during task, in adolescence has not been studied. Therefore, the first aim of this research is to identify how dynamic network flexibility of the striatal-cortical circuits supports reinforcement learning in adolescence (13-17 years). Quantifying the modular architecture of an evolving system over the course of learning in a sample that performs better will provide insight into how the brain is optimized for learning associations in the environment and will give us clues into how these neural circuits develop between adolescence and young adulthood. We predict that greater network flexibility will be associated with better learning performance in an adolescent sample.

2. What is the role of dynamic network flexibility of the hippocampus in learning performance?

Prior research has found that reinforcement learning behavioral benefits in adolescents are related to functional activation of the hippocampus to prediction error as well as greater hippocampus and putamen connectivity (Davidow et al. 2016). Therefore, the second aim of this study is to use network analysis to measure dynamic flexibility of the hippocampus to identify whether dynamic changes in network communication with the hippocampus explain better performance in the adolescent sample. We expect that greater DNF of the hippocampus will predict learning over time and may explain age group differences in better learning performance and reinforcement learning model parameters.

3. Does whole-brain flexibility predict individual learning rate in adolescents?

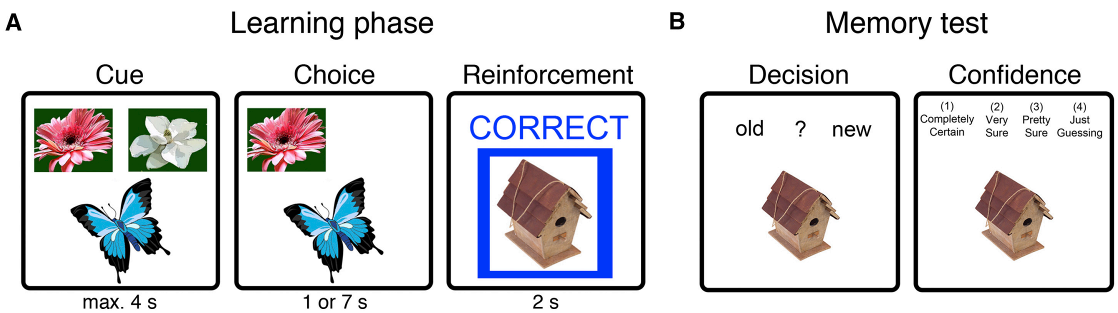
While whole-brain dynamic flexibility during reinforcement learning has been investigated in the adult sample, it has yet to be characterized during adolescence. Given that whole-brain dynamic flexibility predicts a lower learning rate in adults (Gerraty et al. 2018) and adolescents overall have a lower learning rate than adults, we suspect adolescents will have overall greater whole-brain flexibility compared to the adult group. Moreover, we predict that similar to the adult sample, whole-brain dynamic flexibility will be associated with individual difference in learning parameters to predict learning rate over time.

Describe the key dependent variable(s) specifying how they will be measured. (optional)

The key dependent variables in this study are 1. overall performance measured by learning accuracy (percent optimal choice) and 2. estimated parameters from the reinforcement learning model, learning rate (an index of how an individual scales prediction error learning signals to update future choices) and inverse temperature (an index of fidelity to choosing the highest value option, vs exploring the alternative option which may have a lower expected value for the outcome and thus a lower probability of being chosen).

How many and which conditions will participants be assigned to? (optional)

All participants completed the same task in the scanner. 4 probabilistic reinforcement learning blocks followed by a surprise episodic memory test. Stimuli will be counterbalanced across subjects.



Task Design.

Specify exactly which analyses you will conduct to examine the main question/hypothesis. (optional)

Behavioral Analysis

Learning Performance:

Overall performance will be measured by learning accuracy. During each learning phase (4 blocks) a percent correct score will be computed based on optimal choice, regardless of the presented feedback.

Learning Rate:

Based on feedback received and whether the subject made the optimal choice on a trial-by-trial basis we will characterize learning rate using subjects' decisions (choice behavior) using a reinforcement learning model (Daw 2011; Sutton and Barto 1981)). Briefly, this model will evaluate expected value for a given choice at time (t) based on updating of reinforcement outcome via prediction error. With two additional free parameters: 1. alpha - updating, the extent to which value is updated by feedback from a signal trial, and 2. beta - the inverse temperature, the probability of making a particular choice using a softmax function (Daw 2011; Ishii, Yoshida, and Yoshimoto 2002).

fMRI Analysis

Preprocessing

Functional images will be preprocessed using FSL's FMRI Expert Analysis Tool (FEAT (Smith et al. 2004)). Images from each learning block will be high-pass filtered at $f > 0.008$ Hz, spatially smoothed with a 5mm FWHM Guassian kernel, grand-mean scaled, and motion corrected to their median image using an affine transformation with tri-linear interpolation. The first three images will be removed to account for saturation effects. Functional and anatomical images will be skull-stripped using FSL's Brain Extraction Tool. Functional images from each block will be co-registered to subject's anatomical images and non-linearly transformed to a standard template (T1 Montreal Neurological Institute template, voxel dimensions 2mm cubed using FNIRT (Andersson, Smith, and Jenkinson 2008). Following image registration, time courses will be 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 funtional connectivity (Power et al. 2012; Satterthwaite et al. 2012a), time courses will be further proprocessed via a nuisance regression with 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. 2012b).

Dynamic Network Connectivity

Whole Brain

Given how dynamic the brain is in integrating and updating information from the environment, this analysis aims to use a multilayer approach to characterize network modularity and community structure (spatially) over time (temporally) (Jutla, I.S., Jeub, L.G., and Mucha 2011; Mucha et al. 2010). To assess dynamic connectivity of the whole brain we will use 110 subcortical and cortical brain ROIs from the Harvard-Oxford atlas. Time course data will be subdivided into 8 50s time windows of 25 TRs each (per block). For each block, connectivity is quantified as the magnitude-squared coherence between each pair of ROIs at $f = 0.06$-0.12 Hz in order to assess modularity over short time window consistent with previous reports [Bassett et al. (2011); Bassett et al. (2013)). This frequency range is 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. Subjectspecific 110x110x8 connectivity matrices for each learning block (4) will be created containing coherence values ranging from 0 and 1. Regional distribution of network flexibility across the whole brain will be examined across the entire learning task. We predict that whole-brain flexibility (averaged across all ROIs) will increase in early learning blocks and decrease in later stages of the task in adolescence, as found in the adult sample.

To examine the effects of flexibility on learning behavior, we will estimate a generalized mixed-effects model predicting optimally correct choices with flexibility estimates for each block with a logistic link function, using Maximum Likelihood (ML) approximation implemented in the lme4 package (Bates et al. 2015). We will include 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 will also be included as a fixed effect in the model.

Striatal & Hippocampal Flexibility & Learning Performance

We will also probe two a priori seeds, the striatum and hippocampus bilaterally. Flexibility measures, the extent to which a region changes its community allegiance over time (Bassett et al. 2011), will be computed for each learning block. Flexibility of a priori regions will be compared to a generalized flexibility score assessed across the whole-brain for each subject.

Correlations between striatum & hippocampus flexibility and learning performance in adolescents will be run. We will also probe distinct sub-regions of the striatum, we will perform separate tests with the right and left caudate, putamen, and nucleus accumbens. Will we also further look at subsections of the hippocampus (i.e., anterior).

Any secondary analyses? (optional)

Memory Performance & Hippocampal Flexibility

Followed by the scanning session, subjects were given a surprise test of recognition memory for the images that co-occurred with feedback presentations during the learning task.

To explore the relationship between network dynamics and the important role of the hippocampus in learning and memory systems, we will also regress flexibility statistics against subsequent memory scores for the trial-unique objects presented during feedback. Previous reports in adults find no association between memory and the striatal ROI; however, given the role of the hippocampus in episodic memory and the important finding that hippocampus connectivity relates to better learning the adolescent sample, we will conduct an exploratory analysis to examine the role of network flexibility of the hippocampus in memory performance.

How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined. (optional)

The sample will include usable data from 22 adults, 24-30 years old, (total 25 scanned, 3 excluded for two technical issues and one incidental finding) and 25 adolescents, 13-17 years old, (total 28 scanned, two excluded for technical issues and one for excessive motion). Behavioral data were collected from a total of 31 adults and 43 adolescents. These numbers are based on the data quality and how many individuals have already participated in the experiment. There is a possibility additional subjects will be excluded for additional motion correction preprocessing outline above (see preprocessing)

Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?) (optional)

Data exclusion

As previous research on these groups has done, subjects will be excluded for technical issues in behavioral data collection and incidental neurological findings and excessive motion in the fMRI task.

Variables collected for exploratory purposes

1. Prediction Error results

2. Demonstration of learning without reinforcement

3. Resting state data

Subjects participated in two resting state scans, one before the learning blocks and one after the learning blocks. Exploratory analysis may include differences in network flexibility during resting state before and after task. Differences between task and rest network flexibility may also be examined in understanding the system development independently of task stimuli, given that executive network switching has been shown to increase with age (Chai et al. 2017).

4. Subsequent memory test

Followed by the scanning session, subjects were given a surprise test of recognition memory for the outcome images of the learning task.

Bibliography

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

Bassett, Danielle S, Nicholas F Wymbs, Mason A Porter, Peter J Mucha, Jean M Carlson, and Scott T Grafton. 2011. "Dynamic reconfiguration of human brain networks during learning." *Proceedings of the National Academy of Sciences of the United States of America* 108 (18). National Academy of Sciences: 7641–6. doi:10.1073/pnas.1018985108.

Bassett, Danielle S., Nicholas F. Wymbs, M. Puck Rombach, Mason A. Porter, Peter J. Mucha, and Scott T. Grafton. 2013. "Task-Based Core-Periphery Organization of Human Brain Dynamics." Edited by Olaf Sporns. *PLoS Computational Biology* 9 (9). Public Library of Science: e1003171. doi:10.1371/journal.pcbi.1003171.

Bates, Douglas, Martin Mächler, Benjamin M Bolker, and Steven C Walker. 2015. "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software 67.* https://arxiv.org/pdf/1406.5823.pdf.

Casey, B J, Rebecca M Jones, and Todd A Hare. 2008. "The adolescent brain." *Annals of the New York Academy of Sciences* 1124 (March). NIH Public Access: 111–26. doi:10.1196/annals.1440.010.

Chai, Lucy R., Ankit N. Khambhati, Rastko Ciric, Tyler M. Moore, Ruben C. Gur, Raquel E. Gur, Theodore D. Satterthwaite, and Danielle S. Bassett. 2017. "Evolution of brain network dynamics in neurodevelopment." *Network Neuroscience* 1 (1). MIT Press One Rogers Street, Cambridge, MA 02142-1209USAjournals-info@mit.edu: 14–30. doi:10.1162/NETN_a_00001.

Cohen, Jessica R, Robert F Asarnow, Fred W Sabb, Robert M Bilder, Susan Y Bookheimer, Barbara J Knowlton, and Russell A Poldrack. 2010. "A unique adolescent response to reward prediction errors." *Nature Neuroscience* 13 (6). NIH Public Access: 669–71. doi:10.1038/nn.2558.

Damien, A. Fair, Nico, U. F. Dosenbach, Jessica, A. Church, Alexander, L. Cohen, Shefali, Brahmbhatt, Francis, M. Miezin, Deanna, M Barch, Marcus, E. Raichle, Steven, E. Petersen, Bradley, L. Schlaggar. 2007. "Development of distinct control networks through segregation and integration." *PNAS* 104 (33): 13507–12. http://www.pnas.org/content/104/33/13507.full.

Davidow, Juliet Y., Karin Foerde, Adriana Galvan, and Daphna Shohamy. 2016. "An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence." *Neuron*. doi:10.1016/j.neuron.2016.08.031.

Daw, Nathaniel D. 2011. "Trial-by-trial data analysis using computational models." *Decision Making, Affect, and Learning: Attention and Performance XXIII* 23: 3–38. doi:10.1093/acprof:oso/9780199600434.003.0001.

Gerraty, Raphael T, Juliet Y Davidow, Karin Foerde, Adriana Galvan, Danielle S Bassett, and Daphna Shohamy. 2018. “Dynamic flexibility in striatal---cortical circuits supports reinforcement learning.” doi:10.1101/094383.

Gu, Shi, Theodore D Satterthwaite, John D Medaglia, Muzhi Yang, Raquel E Gur, Ruben C Gur, and Danielle S Bassett. 2015. “Emergence of system roles in normative neurodevelopment.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (44). National Academy of Sciences: 13681–6. doi:10.1073/pnas.1502829112.

Ishii, Shin, Wako Yoshida, and Junichiro Yoshimoto. 2002. “Exploitation-exploration control Control of exploitation-exploration meta-parameter in reinforcement learning” 15 (4): 665–87. https://pdfs.semanticscholar.org/1c5e/37e86f2a45d85a5ed96483fe08369e5b7d9e.pdf.

Jutla, I.S., Jeub, L.G., and Mucha, P.J. 2011. “A generalized Louvain method for community detection implemented in MATLAB.” http://netwiki.amath.unc.edu/GenLouvain.

Mucha, Peter J, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. 2010. “Community Structure in Time-Dependent, Multiscale, and Multiplex Networks.” https://arxiv.org/pdf/0911.1824.pdf.

Pagnoni, Giuseppe, Caroline F. Zink, P. Read Montague, and Gregory S. Berns. 2002. “Activity in human ventral striatum locked to errors of reward prediction.” *Nature Neuroscience* 5 (2). Nature Publishing Group: 97–98. doi:10.1038/nn802.

Power, Jonathan D., Kelly A. Barnes, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2012. “Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion.” *NeuroImage* 59 (3). Academic Press: 2142–54. doi:10.1016/J.NEUROIMAGE.2011.10.018.

Satterthwaite, Theodore D., Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, et al. 2012. “An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data.” *NeuroImage* 64 (January). Academic Press: 240–56. doi:10.1016/J.NEUROIMAGE.2012.08.052.

Satterthwaite, Theodore D., Daniel H. Wolf, James Loughead, Kosha Ruparel, Mark A. Elliott, Hakon Hakonarson, Ruben C. Gur, and Raquel E. Gur. 2012. “Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth.” *NeuroImage* 60 (1). Academic Press: 623–32. doi:10.1016/J.NEUROIMAGE.2011.12.063.

Smith, Stephen M., Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E.J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, et al. 2004. “Advances in functional and structural MR image analysis and implementation as FSL.”

NeuroImage 23 (January). Academic Press: S208–S219. doi:10.1016/J.NEUROIMAGE.2004.07.051.

Sutton, Richard S, and Andrew G Barto. 1981. "Toward a Modern Theory of Adaptive Networks: Expectation and Prediction." *Psychological Review* 88 (2): 135–70. http://psycnet.apa.org/fulltext/1981-20731-001.pdf.