Supplementary Materials:

Thalamic regulation of reinforcement learning strategies across prefrontal-striatal networks

Bin A. Wang[1], Mien Brabeeba Wang[2], Norman H. Lam[3], Liu Mengxing[3], Shumei Li[4], Ralf D. Wimmer[3], Pedro M. Paz-Alonso[5,6], Michael M. Halassa[3,7#✉], and Burkhard Pleger[8,9#]

[1]School of Psychology, Center for Studies of Psychological Application, Guangdong Key Laboratory of Mental Health and Cognitive Science, Key Laboratory of Brain Cognition and Educational Science, Philosophy and Social Science Laboratory of Reading and Development in Children and Adolescents, South China Normal University, Guangzhou, China

[2]Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

[3]Department of Neuroscience, Tufts University School of Medicine, Boston, MA, USA

[4]Department of Medical Imaging, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, China

[5]BCBL. Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain

[6]Ikerbasque, Basque Foundation for Science, Bilbao, Spain

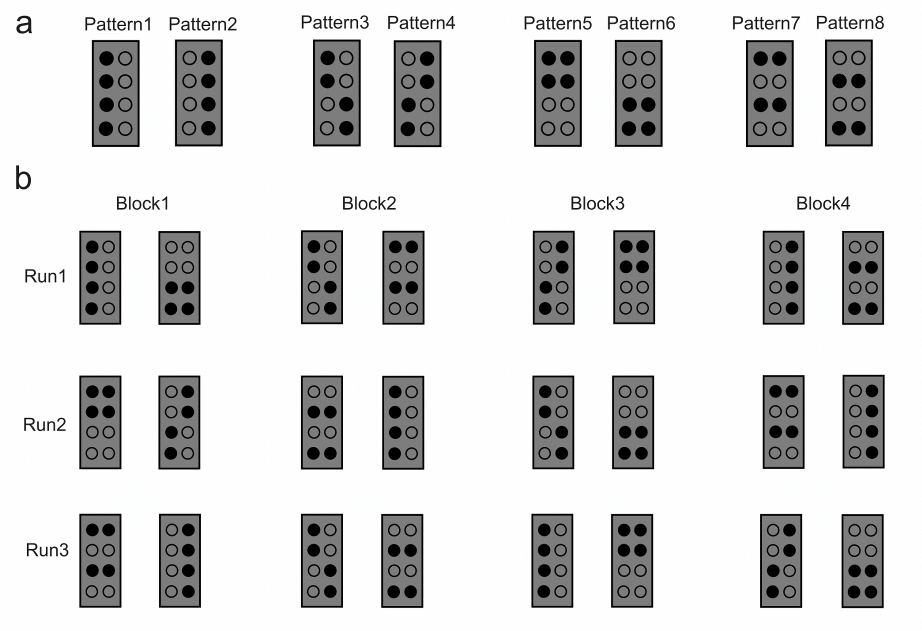
[7]Department of Psychiatry, Tufts University School of Medicine, Boston, MA, USA

[8]Department of Neurology, BG University Hospital Bergmannsheil, Ruhr-University Bochum, Bochum, Germany

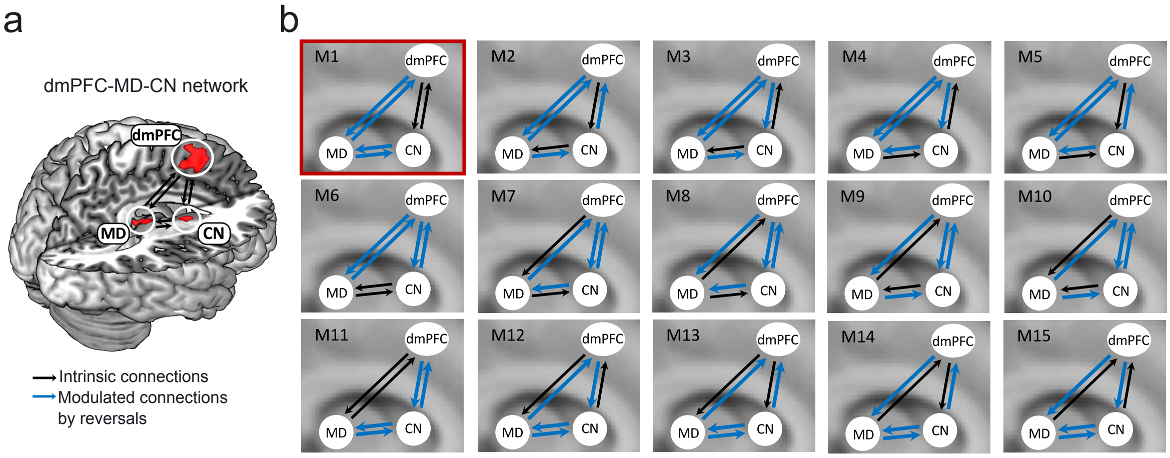
[9]Collaborative Research Centre 874 “Integration and Representation of Sensory Processes”, Ruhr-University Bochum, Bochum, Germany

[#] These authors jointly supervised this work

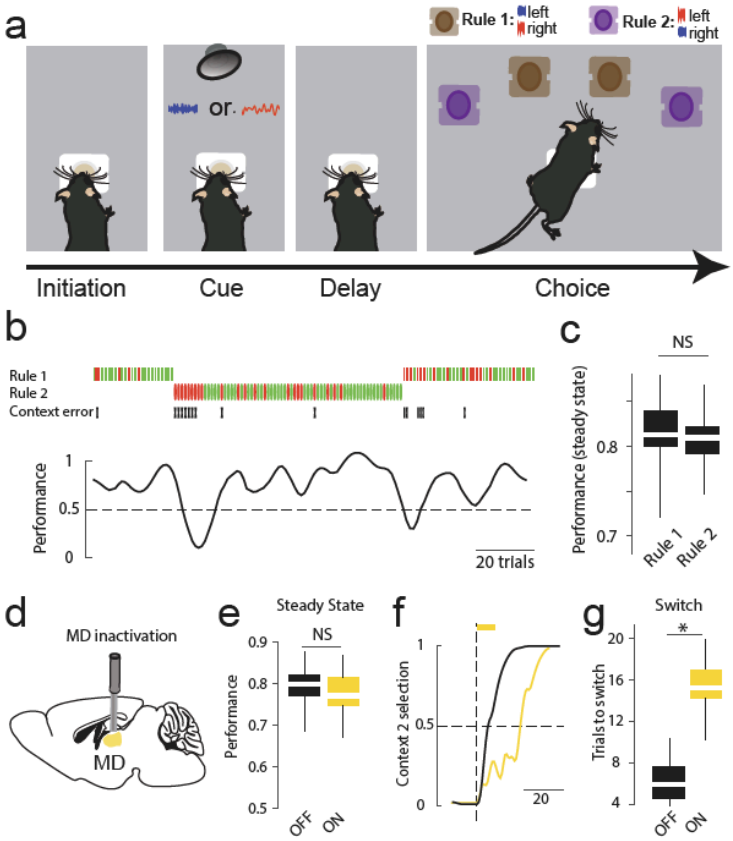
✉ Correspondence to: michael.Halassa@tufts.edu (M.M.H)



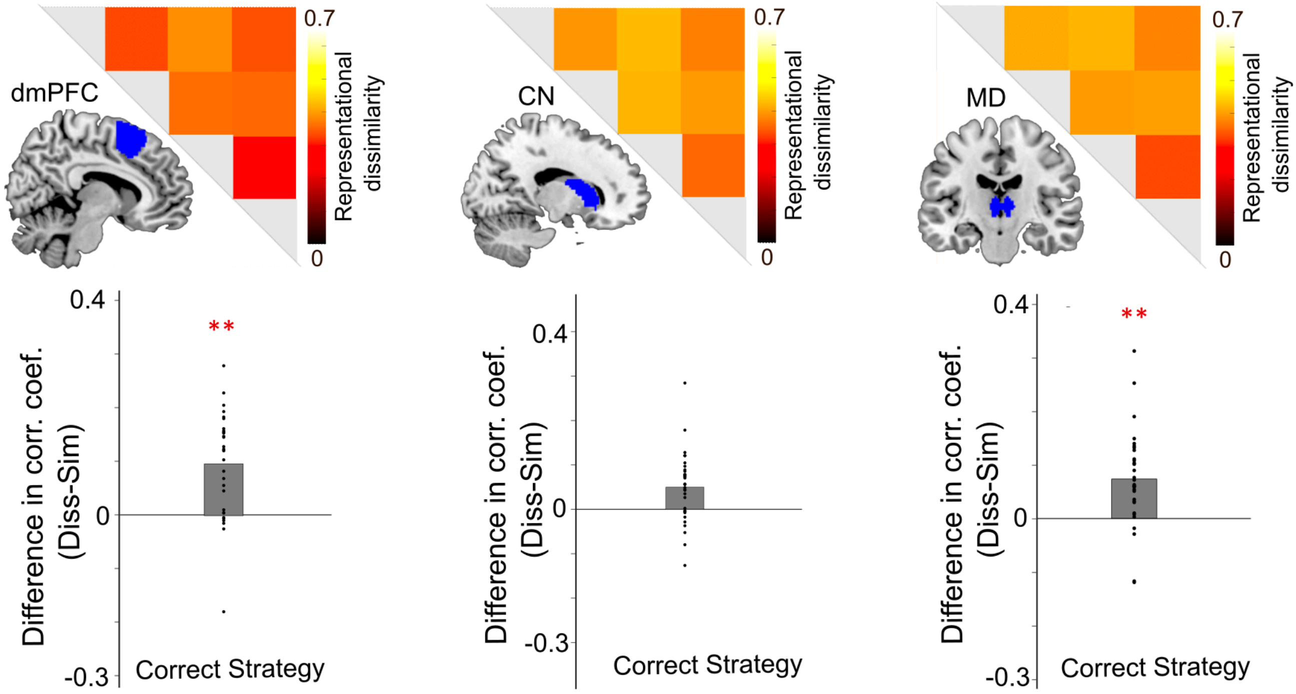
Supplementray Fig.1 The tactile patterns used for each bloch in three runs. a. There were eight alternative patterns in total. Each pattern was characterized by four raised pins, and four lowered pins. **b.** A pair of tactile patterns were randomly selected from eight alternative patterns for each new block.



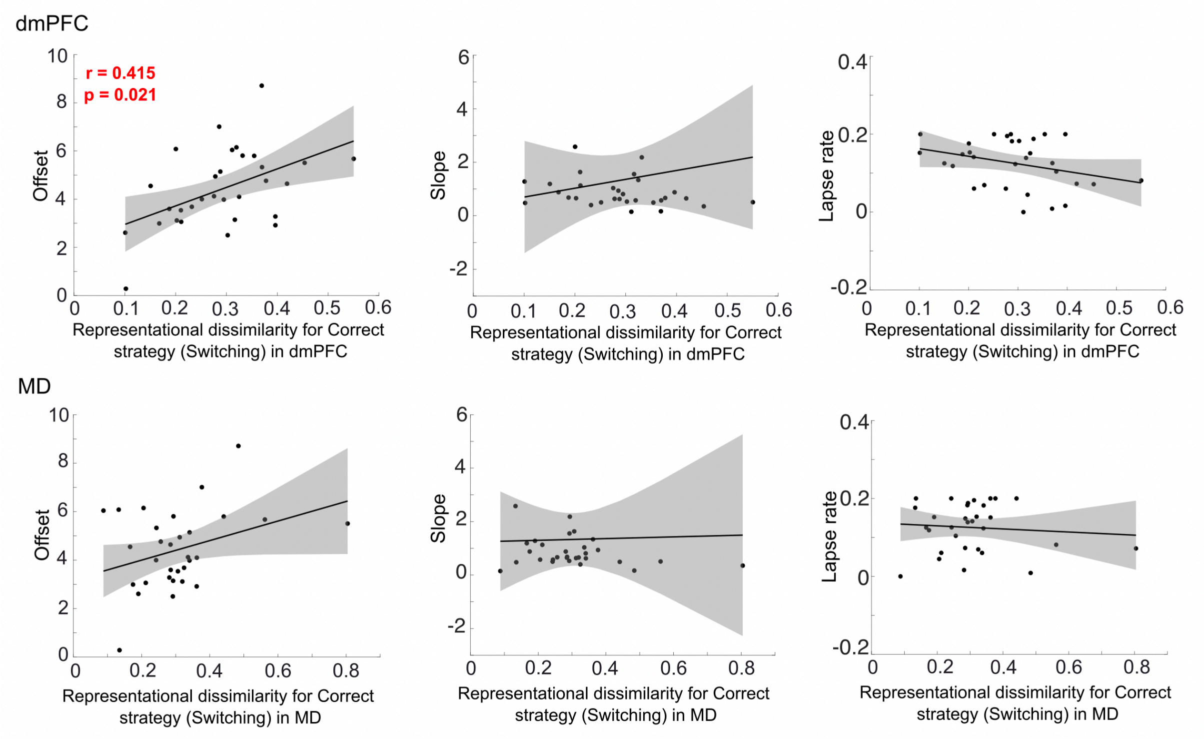
Supplementray Fig.2 The efferent connections from MD to dmPFC and CN were modulated by the rule reversals. a. The dmPFC-MD-CN network with reciprocal endogenous connections among the three regions (DCM A-matrix). The 3D brain was created using MRIcron[1]. The resource of MRIcon was obtained from www.nitrc.org. **b.** Illustration of the model space for Bayesian model selection. First, we specified three model families to determine whether the driving input is either directed to MD, dmPFC, or CN (families not shown in the figure). For each of these three families, we then specified models with different modulatory (bilinear) effects on four connections, which resulted in 15 models. MD-mediodorsal thalamus, dmPFC-dorsomedial prefrontal cortex, and CN-caudate nucleus.



Supplementray Fig.3 Mediodorsal thalamus is required for optimal switching behavior in a comparable rule reversal task in mice. a. Task schematic of the mouse rule reversal task. Mice were cued with a 100ms pulse of high- (blue) or low-pass (red) filtered noise to respond on a left or right response port for liquid reward. The rule (which side a given cue signals) reverses covertly over blocks and animals reported this reversal by responding on a different set of response ports (inner (brown) or outer (purple) pair). **b.** Example performance across two rule reversals. Green and red ticks indicate correct and incorrect responses, respectively. **c.** Animals performed equally well on both rule contexts (n = 17 sessions from two mice, Mann-Whitney U test, U=137; p = 0.81). **d.** Optogenetic silencing of the mediodorsal thalamus (MD). **e**. MD inactivation during the feedback period on trials during *Steady State* had no effect on performance (n = 6 sessions from two mice, U = 22, p = 0.59). **f.** Example behavior without (black) and with (yellow) MD silencing during the feedback period of the first 8 trials after a rule reversal (*Switch*). **g.** Summary of the effect shown in f. Note that MD silencing at the switch significantly delayed switching behavior (n = 7 sessions from two mice, U = 1, p = 0.01).



Supplementray Fig.4 The representation of the decision strategy from the anatomically-defined dmPFC, caudate and MD ROIs. The MD and caudate ROI were defined based on the AAL mask. The dmPFC ROI was defined using the Brainnetome Atlas (Fan et. al., 2016, Cerebral Cortex). The significant representations in both dmPFC and MD for Correct Strategy (one-sided permutation test: dmPFC: effect size = 0.84, p = 0.001, MD: effect size = 0.77, p = 0.002, n=32 participants) were found, but not in the CN (effect size = 0.54, p = 0.029, Bonferroni corrected p < 0.05/2). For Wrong Strategy, none of the three regions exhibited significant representations. The brain slices was created using MRIcron[1]. The resource of MRIcon was obtained from www.nitrc.org. Two asterisks indicate p < 0.01.



Supplementray Fig.5 Correlation between the behavioral parameters and the representational dissimilarity in dmPFC and MD. We found the marginally significant correlation between the representational dissimilarity for *Correct Strategy* (*Switching*) in dmPFC and switch offset (Spearman correlation, $r = 0.415$, $p = 0.021$, 0.05/3 for multiple comparisons, n = 32 participants), suggesting that participants with stronger representations (lower dissimilarity) of the *Correct Strategy* (*Switching*) in the dmPFC switch more quickly after rule reversals.

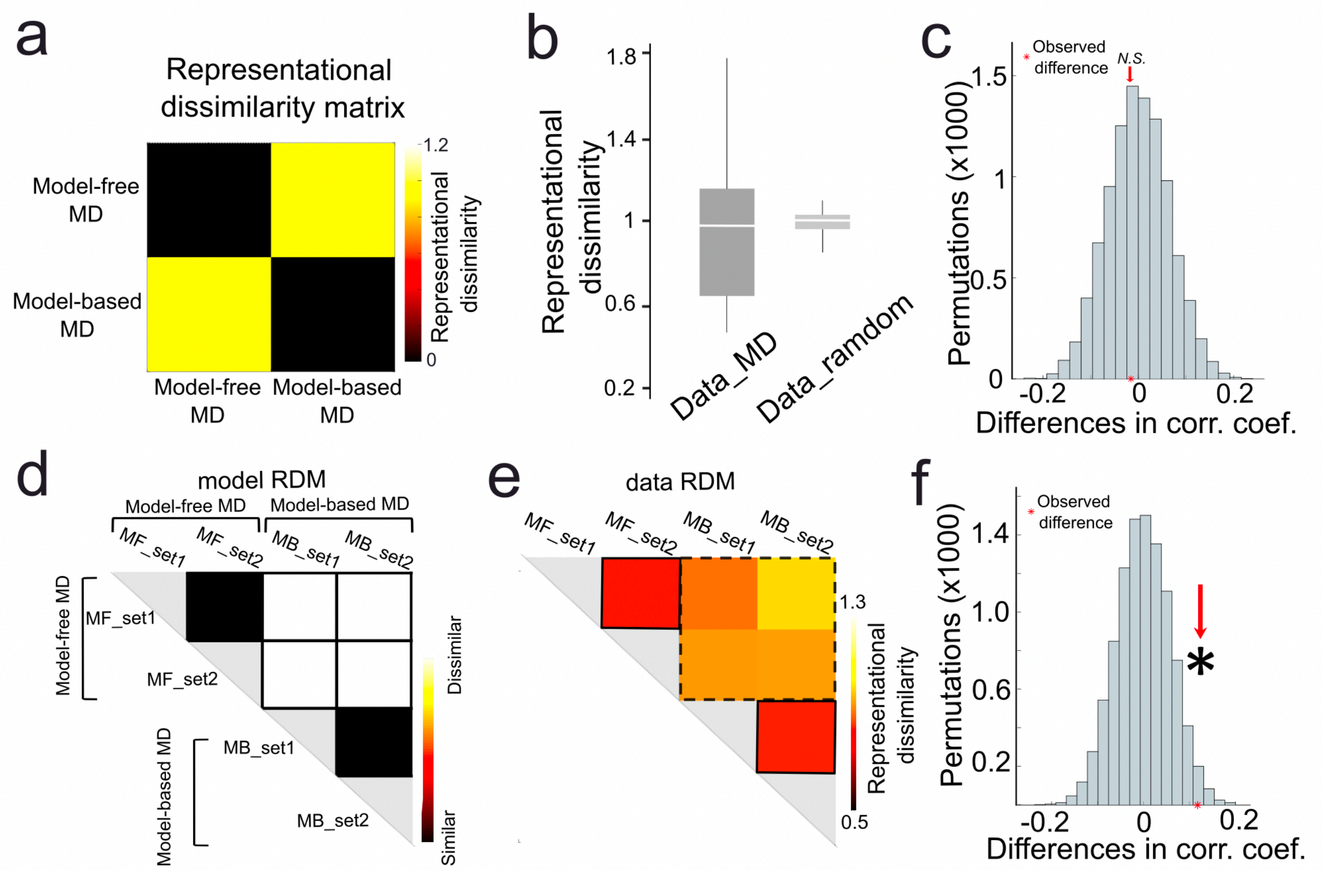
The RSA Analysis for Model-free and Model-based MDs

To confirm the functional separation of MD in model-free and model-based strategy, we perfromed RSA analyses to test whether the spatial response patterns of MD in model-free and model-based blocks are distinct.

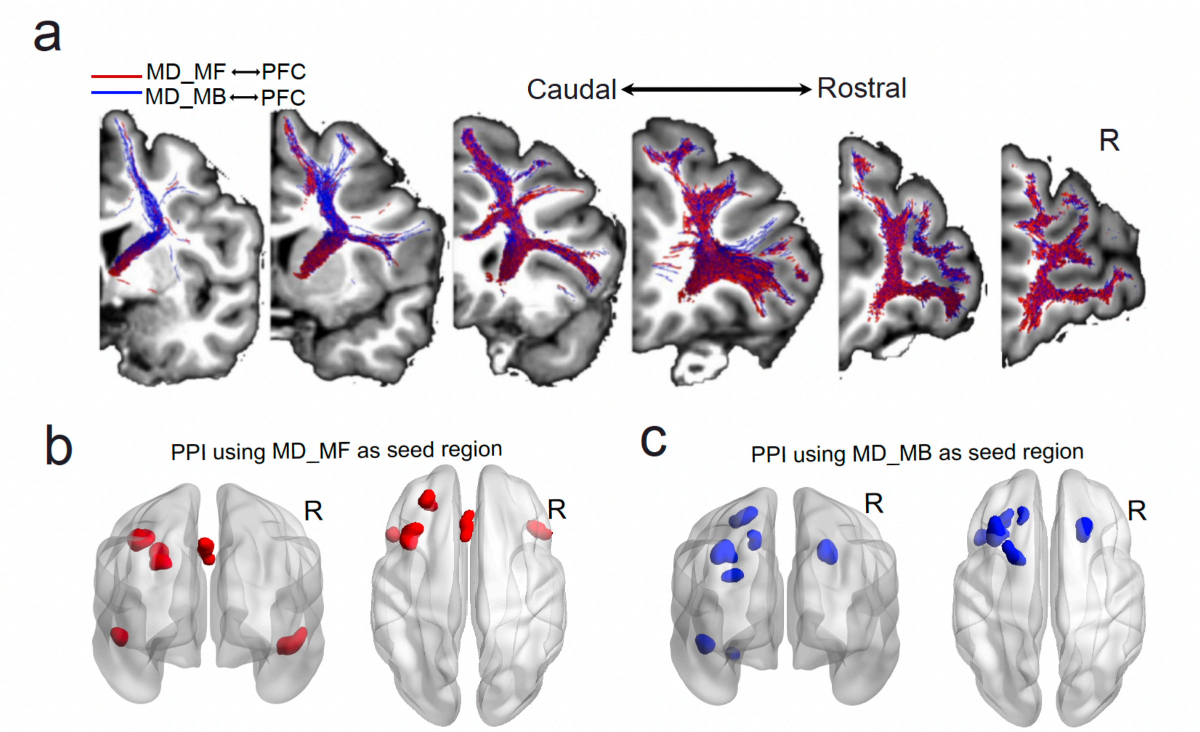
Specifically, we created the MD mask from the contrast of ‘Switch’ vs ‘Steady State’ ($p < 0.001$). Using unsmoothed t-statistic maps, we extracted the activity patterns from the MD seperately for the model-free and model-based blocks. The relative similarity between the response patterns between model-free and model-based blocks was assessed using Pearson correlations and expressed as correlation coefficients (Suppelementary Fig. 6a). To quantify the strength of the similarity between the MD response patterns in model-free and model-based blocks, we compared it to the similarity derived from random data generated by shuffling the sequences of the voxels. Given that random data can be considered as completely unrelated with high dissimilarity, finding that the data from the MDs exhibit a similar level of dissimilarity to that of the random data would indicate that the response patterns in model-free and model-based MD are entirely unrelated. Our results showed no significant difference (Suppelementary Fig. 6b and c) in representational dissimilarity, confirming that the response patterns from model-free and model-based MD are completely different.

Additionally, we hypothesized that there would be low dissimilarity within the MF-related or MB-related MD, but high dissimilarity between MF-related and MB-related MD if the response patterns of model-free and model-based MDs are different. Based on this predicted correlation distance between MF and MB blocks, we constructed the representational dissimilarity matrices (RDMs) of the model (Supplementary Fig. 6d). It is important to note that we divided the MF and MB blocks into two halves (MF_set1/MF_set2, MB_set1/MB_set2), and the response patterns of the MD were extracted from these four contexts. The data RDM from the MD masks is shown in Supplementary Fig. 6e. A permutation test revealed that the dissimilarity between the MD activity patterns from different reinforcement learning strategies is significantly higher (the dashed square elements versus solid square elements in

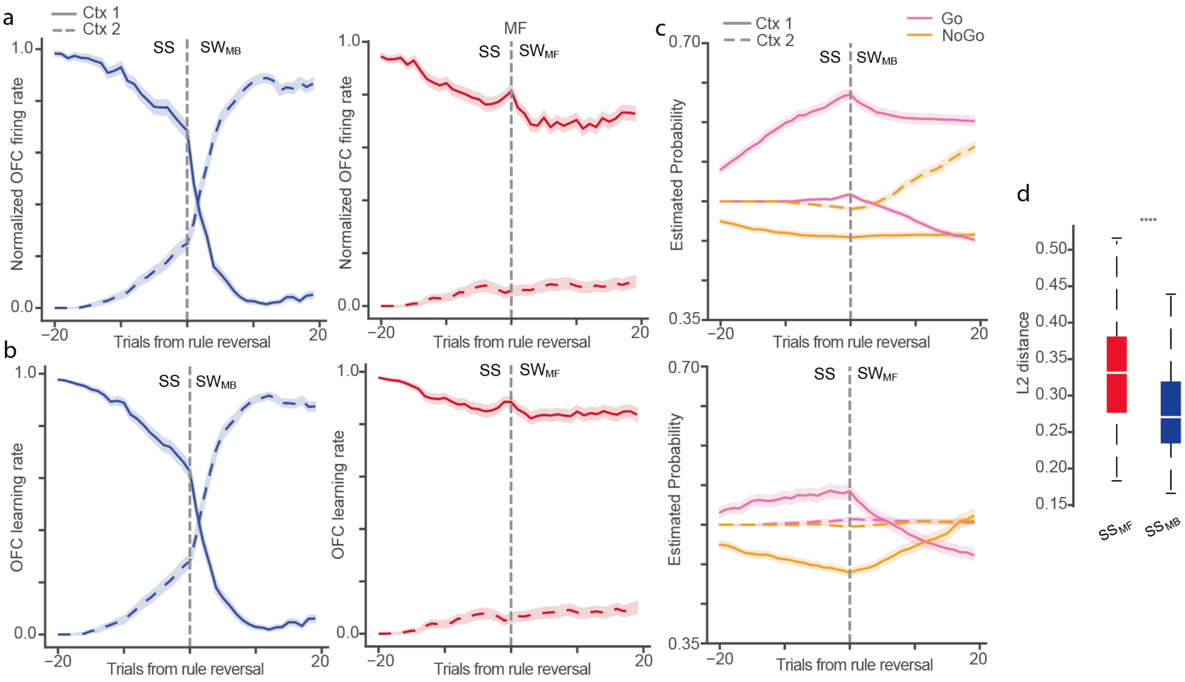
Supplementary Fig. 6e, effect size = 0.52, p = 0.02, Supplementary Fig. 6f), thereby validating the distinct response patterns of model-free and model-based MD.



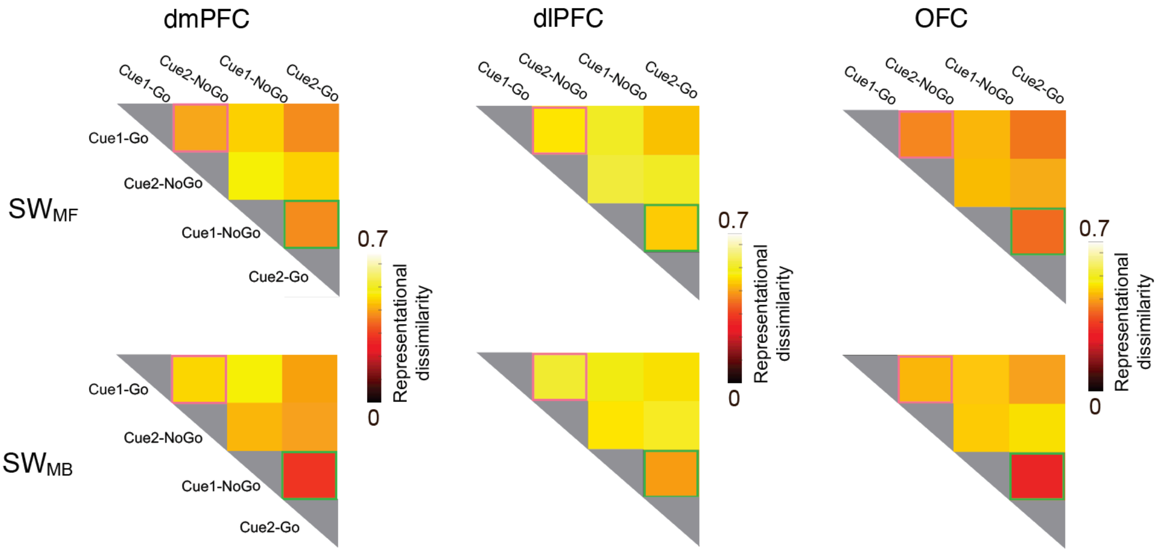
Supplementray Fig.6 The multivariate reprensentational similarity analysis (RSA) indicates that the MD response patterns encoding model-free and model-based reinforcement learning are independent. a. The data reprensentational dissimilarity matrix (RDM) between the model-free MD and model-based MD**.** The activity patterns of model-free and model-based MD were extrated from the unsmoothed t-statistic maps of parametric regressors (RPE and SPE) of GLM of fMRI data. The MD masks were based on the results in Fig. 3d and 3f. The relative dissimilarity between the activity patterns for the model-free and model-based MD was assessed using Pearson correlations and expressed as 1- correlation coefficients. **b.** The box plots of dissimilarities for the data of MD and random shuffling of MD voxels (n=32 participants). Box plots indicate the median (middle line), 25th, and 75th percentile (box), and the maximum and minimum (whiskers). **c**. The dissimilarities between the MDs were compared with that after random shuffling of MD voxels using a two-sided permutation test. The result showed that there was no significant difference of representational dissimilaritiy between the real MD data and shuffled MD data (effect size = -0.06, p = 0.81), suggesting the mutual independence of activity patterns between MD response patterns encoding model-free RPE and model-based SPE. **d.** The validation of RSA analysis by dividing the trials into two halves (MF_set1/MF_set2, MB_set1/MB_set2). We assumed the lower representational dissimilarity between the activity responses from the same reinforcement learning strategy (i.e., MF_set1 vs. MF_set2, and MB_set1 vs. MB_set2, the black elements in model RDM), compared with that between the different reinforcement learning strategy (the white elements in model RDM). **e.** The data RDM from the MD masks. **f.** The permutation test showed that the dissimilarity between the MD activity patterns from different reinforcement learning strategy is significantly higher (two-sided permutation test, effect size = 0.52, *p* = 0.02), indicating that response patterns of model-free and model-based MD are different.



Supplementray Fig.7 The structural and functional connectivity patterns between MDs and PFC regions a. Streamline visualization of the tractography between the MD ROIs and PFC in a representative brain in individual space. Blue: MD(MB)-PFC; Red:MD(MF)-PFC. The 3D brain was created using MRIcron[1]. The resource of MRIcon was obtained from www.nitrc.org. **b.** Visualization of brain regions across both hemispheres characterizing the functional connecitivty with the model-free MD. The figurs were created in BrainNet Viewer[2]. The source of BrainNet Viewer was obtained from www.nitrc.org. **c.** Same with b. but for the model-based MD.



Supplementray Fig.8 Normalized OFC firing rate and learning rate for CogLink. a. Summarized plot (Mean ± SEM) of normalized OFC firing rate at model-based (n = 308) and model-free (n = 192) blocks. **b.** Summarized plot (Mean ± SEM) of normalized OFC learning rate at model-based and model-free blocks. **c**. Summarized plot (Mean ± SEM) of generative model of receiving reward after presenting with cue 1 at model-based and model-free blocks. **d.** A box plot of L2 distance between estimated generative model before switch and true generative model for model-based and model-free blocks (****$p < 10^{-4}$; two-sided rank sum test, MB n = 308, MF n = 192). Box plots indicate the median (middle line), 25th, and 75th percentile (box), and the maximum and minimum (whiskers).



Supplementray Fig.9 Averaged RDMs of response pattern in dmPFC, dlPFC and OFC for model-free (SW_{MF}) and model-based (SW_{MB}) *Swtich* trials respectively. The pink box indicates the *Wrong Strategy* (*Staying*) representation, and the green box the *Correct Strategy* (*Switching*) representation (see Fig.3).

References:

1. Rorden, C. From MRIcro to MRIcron: The evolution of neuroimaging visualization tools. *Neuropsychologia* **207**, 109067 (2025).

2. Xia, M., Wang, J. & He, Y. BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS One* **8**, e68910 (2013).