**Running head: Brain adaptations to volatility**

# Asymmetrical adaptations to increases and decreases in environmental volatility


## Authors

Jie Xu[1,2], Nicholas T. Van Dam[3], Yuejia Luo[1,4-7], André Aleman[2,4], Hui Ai[4*], Pengfei Xu[1,5,8*]

## Affiliations

[1] Faculty of Psychology, Beijing Normal University, Beijing, China

[2] Department of Biomedical Sciences of Cells & Systems, Section Cognitive Neuroscience, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[3] Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Australia

[4] Shenzhen Key Laboratory of Affective and Social Neuroscience, Magnetic Resonance Imaging Center, Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen, China

[5] Center for Neuroimaging, Shenzhen Institute of Neuroscience, Shenzhen, China

[6] Department of Psychology, Southern Medical University, Guangzhou, China

[7] College of Teacher Education, Qilu Normal University, Shandong, China

[8] Great Bay Neuroscience and Technology Research Institute (Hong Kong), Kwun Tong, Hong Kong, China

[*]Corresponding author:
Hui, Ai, Ph.D.
Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen, China
E-mail: hui.ai@hotmail.com

Pengfei Xu, Ph.D.
Faculty of Psychology, Beijing Normal University, Beijing, China.
E-mail: pxu@bnu.edu.cn



## Abstract

Humans adapt their learning strategies to changing environments by estimating the volatility of the reinforcement conditions. Here, we examine how volatility affects learning and the underlying functional brain organizations using a probabilistic reward reversal learning task. We found that the order of conditions was critically important; participants adjusted learning rate going from volatile to stable, but not from stable to volatile, environments. Subjective volatility of the environment was encoded in the striatal reward system and its dynamic connections with the prefrontal control system. Flexibility, which captures the dynamic changes of network modularity in the brain, was higher in the environmental transition from volatile to stable than from stable to volatile. These findings suggest that behavioral adaptations and dynamic brain organizations in transitions between stable and volatile environments are asymmetric, providing critical insights into the way that people learn under uncertainty.


**Keywords:** adaptation, volatility, learning; caudate, dynamic brain network

## Introduction

Learning in an uncertain environment requires flexibility, appropriately adjusting perceived action-outcome associations. Individuals must adjust quickly in dynamic environments to update prior estimation of the association between action and outcome. In stable environments, people relaxedly fine-tune strategies to maintain beliefs about unchanging associations. Flexibly adjusting learning strategies between different environments is critical to optimal performance (*1, 2*). However, adaptations to different environments largely rely on the order in which people experience changes (*3*). People may adapt more quickly in one direction (e.g., moving from a more to less effortful context relative to its opposite). Transfer effects, for example, have historically been shown to favor the difficult to easy direction (*4*).

Uncertainty is an inherent structure of the environment, the estimates of which could be used to characterize organismal adaptability (*5*). Most theoretical propositions of uncertainty are focused on three key types of uncertainty: irreducible uncertainty, estimation uncertainty, and unexpected uncertainty (*5-7*). Irreducible uncertainty, the first-order uncertainty, is represented by risk, wherein the probability of an option-outcome association is known but the outcome of reward or punishment remains uncertain (*8*). Second-order (estimation) uncertainty reflects ambiguity, wherein the probability of a stimulus being associated with a given outcome is unknown and needs to be estimated (*8*). Unexpected (or third-order) uncertainty is volatility, the frequency at which the association between the stimulus and outcome varies dynamically. Irreducible (first-order) and estimation (second-order) uncertainty constitute *expected or known uncertainty* (*5, 7*). In responses to *expected uncertainty*, individuals need more observations to estimate the state of the environment but have to ignore or discount specific surprise outcomes, whereas decisions in *unexpected uncertainty* or volatility rely on the most recent observations given that the association between options and outcomes are likely changing. To adapt to a complex environment, it is necessary to identify and estimate the expected and unexpected uncertainty leading to a surprise event, which may reflect a major change in action-outcome associations. Recent studies suggest that individuals make use of uncertainty to guide their decisions during the learning process (*6, 9, 10*). Difficulties in estimating environmental uncertainty may contribute to maladaptive functions in internalizing psychological disorders (e.g., anxiety, depression) (*11, 12*). Although previous studies have shown different learning rates between stable and volatile environments (*1, 2, 11*), adaptation processes and brain function underlying directional changes in volatility (i.e., the transition from stable to volatile environment vs. from volatile to stable environment) remain unclear.

Adaptation to uncertainty is underpinned by dynamic and distributed brain networks. The anterior cingulate cortex (ACC), a part of the salience network which detects error and conflict, fluctuates co-incidentally with estimated volatility (*1, 2, 9, 13*). Key regions of the default mode network, such as the posterior cingulate cortex (PCC), have also shown to be negatively correlated with unexpected uncertainty (*8, 10, 13*). The orbitofrontal cortex (OFC) and caudate, parts of the reward network, have shown increased activity during changing learning and reward probabilities (*9, 14-17*). One recent study has also shown that surprise and uncertainty during learning are dynamically encoded by the frontoparietal control network, which is linked to

appropriate behavioral adaptation (*18*). Network flexibility, measured by the association of nodes to modules of brain networks, has also been shown to positively predict new learning (*19*). Therefore, the combination of quantitative behavioral models and brain network models, have considerable promise for understanding human learning (*20*).

In this study, we examined how the human brain adapts to transitions in environmental volatility (i.e., the transition from volatile to stable environments in comparison to transitions from stable to volatile environments). Given the historical transfer effect (easier going from hard to easy than vice versa (*4*)), we predicted more effortful adaptation in the direction from volatile to stable environments than the other way around. Dynamic organization of the brain networks underlying the transfer effect associated with environmental volatility were tested using dynamic brain network analyses on functional magnetic resonance imaging (fMRI) data.

## Results

Thirty-seven participants were asked to complete an adjusted probabilistic reward reversal learning task while undergoing functional magnetic resonance imaging (fMRI). On each trial, participants had to choose one of two options with specific reward probabilities (Fig. 1a). In the stable block, two options are consistently associated with reward probabilities of 75% and 25%, respectively (Fig. 1b). In the volatile block, the reward probabilities associated with the options changed between high (80%) and low (20%) every 20 trials (Fig. 1c). To perform optimally, participants had to estimate the reward probabilities of two options from the outcome of previous trials. Option selection and reaction time were recorded. Brain activity was also measured. Three participants were excluded from all analyses as they missed more than 10 of the 180 trials during the primary task (5.6%).

**Behavioral Results.** Learning rate (α) reflects how participants' choice at a given time was influenced by recent previous outcomes. A high learning rate means that current choice is strongly guided by recent outcomes and reflects rapidly changing stimulus-outcome associations. High learning rates are more suitable for volatile environments. In contrast, a low learning rate means that a surprising outcome has little effect on the subsequent choice. Low learning rates are more suitable to stable environments, reflecting that people may not change their selections. We estimated learning rates by fitting a Rescorla-Wagner (RW) learning model to their choices in the three blocks [2, 11]. We also used a Bayesian learning model, which estimates the volatility of the task [2, 11]. Model comparison indicated that the RW model using Grid search provided the best model-to-data fit (RW Grid ML: 115.61 ± 38.57, RW Grid EV: 117.58 ± 37.79, RW fMIN: 122.96 ± 37.88, Bayesian: 260.02 ± 53.82, $F_{(3, 99)} = 139.5$, $p < 0.001$; Fig. 2a). There was no difference between the ML estimate and EV estimate ($p = 0.931$). The RW model with Grid Search also generated the largest correlation between modeled probability and participants' choices (RW Grid ML: 0.801 ± 0.110, RW Grid EV: 0.801 ± 0.108, RW fMIN: 0.771 ± 0.104, Bayesian: 0.736 ± 0.699, $F_{(3, 99)} = 10.615$, $p < 0.001$; Fig. 2b). Therefore, the Bayesian model was only used to estimate environmental volatility.

To understand how participants' behaviors and mental representations changed over time, we examined participants' cumulative true responses and modeled choices. In general, participants could successfully capture changes in probability across the whole

experiment. Compared to the Bayesian model, the RW model fitted behavioral data better in both the stable-volatile-stable task and the volatile-stable-volatile task (Fig. 2c, d). We also estimated the environmental volatility $v_{(i)}$ by using the Bayesian model across the whole task. The estimated volatility increased suddenly after reversal and decreased gradually in the stable state (Fig. 2e, f). However, this variability in the volatile-stable-volatile task was not strong, especially in the stable state, which suggests that participants perceived/expected the environment as having high uncertainty. A repeated measures ANOVA of the Order and Volatility showed different tendencies of estimated volatility between the stable-volatile-stable task and volatile-stable-volatile task ($F_{(1, 32)} = 41.936$, $p < 0.001$, $\eta^2 = 0.567$; Fig. S1). *Post hoc* analyses showed that estimated volatility differed as a function of an interaction with order; the volatile state (0.809) was significantly larger than the stable state (0.584) within the stable-volatile-stable task ($F_{(1, 32)} = 40.85$, $p < 0.001$), whereas the stable state (0.809) was significantly higher than the volatile state (0.712) within the volatile-stable-volatile task ($F_{(1, 32)} = 7.65$, $p = 0.009$).

We observed a decrease in learning when shifting from volatile to stable conditions across both ordered tasks (Fig. 2g). To confirm whether participants adapted their learning rates by order of transition (from stable to volatile state *vs.* from volatile to stable state), adapted learning rates were calculated by subtracting learning rate in the stable state from the volatile state. A repeated measures ANOVA of the Order and Transition showed that the adapted learning rates were significantly higher in volatile to stable transitions (adapted $\alpha_{v\text{-}s} = 0.076$) than stable to volatile transitions (adapted $\alpha_{s\text{-}v} = -0.002$), regardless of the task type ($F_{(1, 32)} = 5.141$, $p = 0.03$, $\eta^2 = 0.138$; Fig. 2h). Further analysis showed that adapted learning rate from volatile to stable state is significantly larger than zero which means that learning rate in the volatile state (0.4514) was larger than in the stable state (0.3714; $t_{(1, 33)} = 2.413$, $p = 0.022$). These results indicate that participants adapted their learning more in the transition from volatile to stable states than the other way around. The learning rate was not significantly different in the volatile environment from the stable environment ($p > 0.05$; Fig. S2). The adapted leaning rate in volatile to stable transitions was not significantly correlated to anxiety, depression or impulsivity ($ps > 0.05$; Fig. S3). Potential correlations were not explored for adapted learning rate in stable to volatile transitions due to lack of effect.

**Neuroimaging Results.** The estimation of environmental volatility in the Bayesian model was negatively associated with activity in the bilateral caudate nuclei (left caudate, peak at x, y, z = -18, 12, 15, 87 voxels; right caudate, peak at x, y, z = 21, 15, 18, 54 voxels; $p_{FWE} < 0.05$; Fig. 3a). The increased blood oxygenation level-dependent (BOLD) signal in the bilateral caudate nuclei reflected lower estimated environmental volatility. Activity of the bilateral caudate nuclei in response to volatility was also associated with the variance of learning rates across the three blocks ($r_{(33)} = 0.437$, $p = 0.01$; Fig. 3b). These results revealed that encoding of environmental volatility in the caudate could predict the adjustment of learning strategies across the three blocks.

To explore brain network configurations related to volatility, we performed PPI analysis, using the bilateral caudate nuclei as seeds. We found that the bilateral caudate nuclei showed positive connectivity with the right dorsolateral prefrontal cortex (dlPFC) and left fusiform gyrus (FFG) related to volatility (dlPFC, peak at x, y, z = -36, 12, 30, 93

voxels; FFG, peak at x, y, z = -18, -90, -6, 118 voxels; $p_{FWE} < 0.05$; Fig. 3c). Correlation analyses showed that functional connectivity between the bilateral caudate nuclei and right dlPFC was positively associated with averaged learning rate across the three blocks ($r_{(33)} = 0.36$, $p < 0.05$; Fig. 3d). Connectivity between the caudate and middle frontal gyrus (MFG) differed as a function of Order and Transition ($F_{(2, 64)} = 3.587$, $p = 0.033$, $\eta^2 = 0.101$; See Fig. 4b). *Post hoc* analysis revealed a significant decrease of caudate-MFG connectivity in the transition from volatile to stable relative to the stable to volatile transition, only for the stable-volatile-stable order ($F_{(1, 32)} = 5.470$, $p < 0.05$; Fig. 4c). To explore whether the caudate-MFG pathway tracked with volatility, we calculated correlations between environmental volatility and connectivity of the bilateral caudate and right MFG in each trial of each participant (Fig. S4). A one-sample *t*-test showed that correlations were not significantly different from zero (*ps*>0.05).

Given that the brain dynamically adapts to the changing environment as a complex system, we examined dynamic modular structures of multilayer brain networks over multiple temporal scales. The modularity index was adapted to identify at which temporal scale the node was best partitioned into communities (Fig. 5a). Modularity analyses showed that $Q$ at the smallest temporal scale (10s, one trial) was significantly higher than the other temporal scales (*ps* < 0.001; Fig. 5b). To examine changes of the community properties among the three blocks, we calculated the network flexibility within each block at 10s intervals (each block comprised 60 time-windows). Network flexibility decreased as a function of learning, especially from volatile to stable transitions (Fig. 5d). Repeated measures ANOVA of Order and Transition showed that adapted flexibility was significantly higher from volatile to stable transitions (adapted flexibility = 0.0054) than from stable to volatile transitions (adapted flexibility = -0.0019), regardless of the task type ($F_{(1, 32)} = 4.672$, $p = 0.038$, $\eta^2 = 0.127$; Fig. 5e). *Post hoc* analysis showed that adapted flexibility from volatile to stable transitions was not significantly larger than zero ($t_{(1, 33)} = 1.589$, $p > 0.05$). The contrast between volatile-to-stable and stable-to-volatile transitions showed that adapted flexibility was associated with activity in the left frontal pole (FP), right superior parietal lobule (SPL), left supramarginal gyrus (SMG), lateral occipital cortex, right OFC, left parahippocampal gyrus and brain stem (*ps* < 0.05; Fig. 5f). Correlation analyses didn't show any significant relationships between adapted flexibility and adapted learning rates (*ps* > 0.05).

## Discussion

The ability to adapt, on multiple levels, to constantly changing environments ensures optimal behavior for reward maximization and punishment minimization. Our results show distinctive behavioral adaptations and brain organizations in transitions between environments with varied volatility. Participants exhibited more adaptation and perceived higher volatility in stable environments among transitions from volatile to stable relative to stable to volatile environments. Subjective volatility of the environment was encoded in the bilateral caudate nuclei and its connectivity with the right dlPFC. Notably, modular organizations of dynamic caudate-dlPFC connectivity changed most within shorter time windows, reflecting the need for the brain to adapt quickly while learning of rapid environmental changes.

The asymmetrical adaptation between the volatile-stable and the stable-volatile transitions, suggest a greater need for adaptation under stable to volatile transitions. Heightened adaptation can be largely explained by early learning among major environmental changes. Participants exhibited greater adaptation under stable to volatile transitions when the first block was stable relative to when the first block was volatile. Asymmetrical adaptation is consistent with previous findings of transfer effects; shifts from a difficult task to an easier one is often more effortful than transfer in the opposite direction (*4*). While previous studies have shown higher learning rates in the volatile environment and in the stable environment (*2, 11*), our findings suggest important carryover effects such that experience with volatile environments may impact subsequent learning in a stable environment and vice versa. The idea has been supported in principle by a recent social norm violation study, which shows that participants preconditioned on unfair offers rejected comparable fair offers less frequently than participants preconditioned on generous offers (*21*).

Several brain systems play key roles in adaptation to environmental transitions. The caudate nucleus, a critically part of the midbrain dopaminergic system, plays a key role in processing incentive salience (*22*)and learning associations between stimuli and responses (*17*). A large number of studies have shown that the caudate encodes several learning parameters, including expected value (*23*), action values (*24*), reward prediction errors (*25, 26*), and learning rates (*16*). A primate study has shown that monkeys with caudate dopaminergic depletions exhibit marked difficulty in reconstructing the stimulus-reward associations after reversal (*15*). The current findings show that the caudate also encodes estimated volatility, suggesting a new role of the caudate in learning. The present result is consistent with previous findings that the caudate is engaged in volatility estimates to fine-tune the weights of recent or remote prediction errors in predicting forthcoming needs for control (*27*). Consistent with previous findings on the direct engagement of the dlPFC in encoding volatility (*9, 14*), our results show a modulatory role of the connectivity between the caudate and right dlPFC in estimating volatility. Accumulating evidence has shown that the frontostriatal network is modulated by cognitive control (*28-30*). Therefore, changes in the frontostriatal pathway associated with cognitive control might be involved in estimates of environmental volatility. Contrary to previous work (*1, 2, 9, 31*), we did not find robust volatility-related signals in the ACC. One potential explanation might be that the magnitude of outcome in our task was fixed rather than variable as in previous studies (*2, 11*). Changing outcome probabilities and outcome magnitudes may have increased task difficulty in prior work. It is possible that strong contrast effects (e.g. high probability with low magnitude vs. low probability with high magnitude) might drive ACC activity in previous studies.

In transition from the volatile-stable direction to the stable-volatile direction, participants showed higher network flexibility, suggesting a dynamic reorganization of the brain in adaptation to environmental volatility. Dynamics of functional brain connectivity has been widely used to examine the complex human cognition of moment-by-moment changes in learning (*32-34*), by providing dynamic measurements of flexibility in coordination among different brain states in responses to adaptative behaviors (*35, 36*). Based on dynamic functional connectivity, modular structures aggregated by small subsystems or modules might facilitate behavioral adaptation (*19*).

These findings point to a pivotal role of the frontoparietal control network in the adapted flexibility of individuals in transitions between environments with different volatility.

In conclusion, to the best of our knowledge, the current work is the first to examine the way that the human brain adapts to transitions in environmental volatility. Our results show asymmetrical behavioral and neural adaptations during the environmental transitions with superior adaptation under transitions from stable to volatile environments rather than the opposite. These flexible adaptations are modulated by the striatal reward system and its dynamic connections with the prefrontal control system. The current work sheds light on the nature of the human brain adaptations to navigation of variable environments in daily life.

## Materials and Methods

**Participants.** Thirty-Seven (20 females) Chinese participants, aged between 18 and 30 years (mean±SD = 21.62±2.79 years), without any history of psychiatric disorders were recruited from several Universities in Beijing. Three participants were excluded from all analyses as they missed more than 10 of the 180 trials during the primary task (5.6%). The study protocol was approved by the local Ethics Committee. All participants provided informed consent.

**General procedure.** To ensure participants understood the probability and reversal components underlying the task, a three-stage training task was implemented (*37*). Participants then completed an adjusted probabilistic reward reversal learning task while undergoing fMRI. Participants completed the trait subscale of the Spielberger's State-Trait Anxiety Inventory (*38*), the Self-rating Depression scale (*39*) and Barrett impulsivity scale (*40*). After the experiment, all participants were fully debriefed and received payment based on their task performance.

**Task design.** A probabilistic reward reversal learning task (Fig. 1a) was adapted from previous reversal learning tasks used to examine learning strategies in environmental volatility (*2*, *11*). The task consists of three blocks of 60 trials in which participants were required to make choices between two options with specific reward probabilities (Fig. 1b, 1c). In the stable block, two options were stably associated with reward probabilities at 75% and 25%, respectively. In the volatile block, the probabilities of the two options switched between high (80%) and low (20%) reward probability every 20 trials. To test whether adaptations differ between the transition from stable to volatile and the transition from volatile to stable blocks, participants were randomly assigned to complete the three blocks either in the order of stable-volatile-stable or volatile-stable-volatile without taking a break. None of the participants was informed that tasks would be divided into three different blocks.

In each trial, two options with horizontal and vertical gratings were presented with a visual angle of approximately 8°. Participants were required to make decisions within 3s. Once they responded, the selected option would be highlighted via a white frame to acknowledge the choice with a duration of 0.2s. Sequentially, a question mark was presented at the center of the screen with a jitter interval at 2-4s, to indicate that the outcome was pending. At the phase of outcome, reward (green "+ 1") or no reward (red "+ 0") was presented at the center of the screen for two seconds. At the end of trial, an inter-trial interval with a fixation cross was presented for 0.8-5.8s, to ensure total trial duration was 10s.

**The Rescorla-Wagner model.** Participants' learning rates in three blocks were estimated by a simple Rescorla-Wagner (RW) learning model (*41*). In $(i+1)^{th}$ trial, the predicted reward probability of option A, $V_{a^{(i+1)}}$, was updated using the following equation,

$$V_{a^{(i+1)}} = V_{a^{(i)}} + \alpha \times \delta_{(i)} \quad (1).$$

The $V_{a^{(i+1)}}$ and $V_{a^{(i)}}$ represent the expected values. Specifically, the predicted reward probability of the chosen option A for the $(i+1)^{th}$ and $i^{th}$ trial. The $\alpha \in [0,1]$ indicates the learning rate and $\delta_{(i)}$ is the prediction error on the $i^{th}$ trial which was calculated by comparing the actual reward $r_{(i)}$ with the expected value $V_{a^{(i)}}$,

$$\delta_{(i)} = r_{(i)} - V_{a^{(i)}} \quad (2).$$

The likelihood of the option chosen by participants on the $i^{th}$ trial was estimated by a sigmoidal probability distribution,

$$P_{a^{(i)}} = \frac{e^{\beta V_{a^{(i)}}}}{e^{\beta V_{a^{(i)}}} + e^{\beta V_{b^{(i)}}}} = \frac{1}{1+e^{\beta\left(V_{b^{(i)}} - V_{a^{(i)}}\right)}} \quad (3).$$

Here, $\beta \in (0,10]$ indicates the inverse temperature parameter which controls the degree to which an option would be chosen. When $V_{a^{(i)}}$ is larger than $V_{b^{(i)}}$, the larger the $\beta$, the higher probability that option A would be chosen.

Participants were instructed that the difference between the probabilities of two options would be clear without mentioning that they were opposite, and both add to one. Although it is theoretically possible for participants to learn about the probabilities of two options independently, the probability of the unchosen option B was calculated as follows,

$$V_{b^{(i)}} = 1 - V_{a^{(i)}} \quad (4).$$

The probability of the second option was fixed to make learning simpler.

**The Bayesian learner model.** We also assume that participants would track the probabilities of option and outcome optimally by a Bayesian rule which has been described in previous studies (*1*, *2*, *11*). Here, we provide a brief overview. In each trial, the outcome $y_{(i)}$, reward or not, was determined by the underlying estimated probability of option $r_{(i)}$:

$$\mathrm{p}(y_i | r_i) \quad (5).$$

The probability of option in the $i^{th}$ trial, $r_{(i)}$, was determined by the probability of the option on the $(i-1)^{th}$, $r_{(i-1)}$, and the estimated volatility on the $i^{th}$, $v_{(i)}$:

$$\mathrm{p}\left(r_i \middle| r_{(i-1)}, v_{(i)}\right) \quad (6).$$

The volatility *v* refers to an estimate of the expected rate of change of *r*. According to the Bayesian model, a low volatility would be estimated and each new outcome would have little influence on the estimate of *r* in a stable environment, whereas a high volatility would be estimated in a fast-changing environment, implying *r* may be expected to change quickly. The model also assumes that the way participants track environmental volatility is same as the way participants track the changing probability,

$$\mathrm{p}\left(v_i \middle| v_{(i-1)}, k_{(i)}\right) \quad (7).$$

The environmental volatility on the $i^{th}$, $v$, was determined by the volatility of the preceding trial, $v_{(i-1)}$, and the control parameter, $k_{(i)}$. A large k implies that stable or volatile environment switches to the other one frequently.

All parameters representing the participant' expectations of the statistics of the environment were modeled by the joint probability distribution,

$$p(r_{i+1}, v_{i+1}, k|y_{\leq i+1}) \propto p(y_{i+1}|r_{i+1}) \int [\int p(r_i, v_i, k|y_{\leq i}) p(v_{i+1}|v_i, k) dv_i] p(r_{i+1}|r_i, v_{i+1}) dr_i \quad (8).$$

**Model fitting.** The learning rates and inverse temperature parameters were firstly fitted by a grid search. However, while increasing the precision of two parameters (e.g., $\alpha$ [0:0.01:1] and $\beta$ (0:0.1:10]), the number of all possible combinations of two parameters across three blocks via grid search requires a very large amount of memory. To resolve this issue, we first adopted the solution that these parameters would be estimated separately for each of the three blocks by the grid search. The first prediction values in block 2 and block 3 (trial 61 and 121) were from previous estimates (trial 60 and 120). We also adopted the function fMINSEARCH (fMIN) in Matlab to estimate the parameters without having to measure every single point. To determine the best-fitting values of the learning rates and inverse temperature parameters, we used both the maximum likelihood estimate (ML) and expected value estimate (EV) (multiplying the value of each bin with the probability of that value). Additionally, trials with no response and trials in which reaction time less than 200ms were excluded.

**Model selection.** The Bayesian information criterion (BIC) was used to assess which model best captured participants' choices. Lower value indicates the better fitting. *L* is the likelihood, *m* is the number of estimated parameters, *N* is the number of observations. The Bayesian observer model includes 3 parameters (*r*, *v*, *k*). The RW model with fMINSEARCH that can look for minima without every single point includes 6 parameters ($\alpha, \beta$ per block). The RW model with Grid search includes 2 parameters. Since this model fitted the participants' choices separately for three blocks, the BIC of each block was first calculated and then added together to get the total BIC for each participant.

$$BIC = -2 \log L + m \log N \quad (9).$$

To examine the fitting effect of model, we calculated the correlation of the participants' choices with probabilities estimated by the RW and Bayesian model. The average correlation coefficient was then used as a corroboration to indicate which model was best.

**fMRI data acquisition.** MRI data were collected on a 3T Siemens Prisma MRI scanner with a 64-channel head-neck coil at the Center for MRI Research, Peking University. Functional MRI images were acquired with a simultaneous multiband echo planar imaging (EPI) sequence (TR/TE = 1000/30 ms; FOV = 224 × 224 mm; matrix = 64 × 64; slice thickness = 3.5 mm; slice number = 34; flip angle = 73°; multiband factor = 2). High spatial resolution T1-weighted anatomical images were obtained with the magnetization□prepared rapid gradient□echo (MPRAGE) sequence (TR/TE = 2530/2.98 ms; FOV = 256 × 256 mm; matrix = 256 × 256; slice thickness = 1 mm; flip angle = 7°).

**fMRI data preprocess.** All images were preprocessed using SPM12 (Wellcome Trust Center for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The fMRI data were first corrected for head motion, and the realigned images were coregistered to the T1 structural image which had been manual reoriented to the anterior commissure and then segmented into white matter, gray matter, cerebrospinal fluid (CSF), bone, soft tissues, and air using default tissue probability maps of SPM12. The images were then normalized to the standard Montreal Neurological Institute (MNI)

space with final resolution of 3 $mm^3$. Finally, normalized images were smoothed with a Gaussian kernel of 6 mm full-width at half-maximum.

**fMRI data analysis.** To identify the brain responses to the environmental volatility, we constructed a general linear model (GLM) with the onsets of the options presentation modulated by a parametric regressor (the environmental volatility), onsets of the choice selection, onsets of the jitter period modulated by the volatility, onsets of the feedback delivery modulated by the volatility, onsets of trials with insufficient response time (< 200 ms, including non-response trials), onsets of trials with large head motion (the framewise displacement, FD > 0.5). Additionally, six estimated head movement regressors and a FD regressor were also included as covariates of no interest. The regressors in the GLM design matrix were then convolved with the canonical hemodynamic response function (HRF). Group analyses for brain activation were performed with a random-effect model using a one-sample *t*-test. All results were whole-brain corrected for multiple comparison by a voxel-wise uncorrected threshold at $p < 0.001$ with a family-wise error (FWE) corrected for cluster-level at $p < 0.05$.

**Psychophysiological interaction analyses.** To identify the volatility-specific changes in the interaction between brain regions in whole brain functional connectivity, the psychophysiological interaction (PPI) was performed by using the significant cluster as seed region of interest (ROI) which was related with estimated volatility. For each participant, the first eigenvariate of the ROI was extracted to get the individual voxel time-course. In order to generate the PPI interaction term, this time-course was deconvolved with the canonical HRF and then multiplied by the vector of the estimated volatility. This interaction term was then convolved with the canonical HRF and entered into a PPI GLM along with the vectors of the onsets for the estimated volatility during the feedback, the original eigenvariate time-course and covariates of no interest (six head movement and a FD). After that, we performed second-level analyses with one-sample *t*-test for the contrast images of the PPI interaction term. The results were corrected for a voxel-wise uncorrected thresholded at $p < 0.001$ with an extent FWE-corrected cluster-level $p < 0.05$.

**Dynamic Functional connectivity.** To examine dynamic changes between the seed-ROI and the significant clusters from PPI analysis, we extracted BOLD signals from the corresponding ROIs based on the AAL atlas. The linear Pearson's correlations between the seed and target regions were calculated in three blocks to estimate individual dynamic functional connectivity.

**Dynamic Brain Networks.** To explore modular organizations of the brain in the dynamic adaptation, the modular structures spanning several temporal scales were constructed according to previous recommendations (*19*). The smoothed images were temporally detrended to reduce the effects of linear drift and nuisance signals were removed to reduce the effects of non-neuronal fluctuation, including head motion, the white matter and the CSF. fMRI data were then band-pass filtered to reduce the effects of low frequency drift and high-frequency physiological noises with 0.06-0.12 Hz (*19*). The whole brain was parcellated into 112 ROIs identified in the 3mm Harvard-Oxford (HO) atlas. To construct the individual functional connectivity matrices, the mean BOLD time series were first estimated by averaging voxel time series in each ROI and the linear Pearson's correlation $r_{ij}$ between all pairs of ROIs $i$ and $j$. To correct for multiple comparisons, we first computed the *p*-values $p_{ij}$ of each $r_{ij}$ using the

MATLAB function *corrcoef* and then tested the significance of $p_{ij}$ using a False Discovery Rate (FDR) of $p < 0.05$. The $p_{ij}$ of the correlation matrix elements $r_{ij}$ which passed the FDR corrections was retained. Otherwise, the correlation matrix elements $r_{ij}$ were set to zero. These corrected matrices $r'_{ij}$ constituted adjacency matrices **A,** elements of which $A_{ij} = r'_{ij}$. Based on our experimental setting, we measured functional connectivity over several temporal windows: [10s, 20s 30s, 40s, 50s, 60s, 75s, 100s, 120s, 150s, 200s, 300s, 600s, 1800s]. The node in each functional connectivity matrix was partitioned into a community by maximizing the modularity index $Q$ (*42*). To identify organizations of subtle networks, we constructed the undirected weighted graphs that preserving the information of the strength of connections $r_{ij}$ (*43*). A spectral optimization algorithm was used to optimize the modularity index $Q$ *(44)*. The most popular formula of $Q$ was used (*44*) as follows,

$$Q = \frac{1}{4m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta\left(g_i, g_j\right) \qquad (10),$$

where $m = \frac{1}{2}\sum_{ij} A_{ij}$, $k_i$ is the strength of node $i$, $k_j$ is the strength of node $j$. When node $i$ and node $j$ are in the same module, $\delta\left(g_i, g_j\right) = 1$; Otherwise, it equals 0. Both positive and negative weighted correlation coefficients were used to construct the community with the assumption that it would provide more useful information about the modularity partitions than only using positive correlation matrix elements (*45*). The modularity was generalized (*46*) as

$$Q^{\pm} = \frac{1}{w^+ + w^-}\sum_{ij}\left[A_{ij} - \left(\gamma^+ \frac{w_i^+ w_j^+}{w^+} - \gamma^- \frac{w_i^- w_j^-}{w^-}\right)\right]\delta\left(g_i, g_j\right) \qquad (11),$$

where $\gamma^+$ and $\gamma^-$ are resolution parameters which usually set as one for simplicity and $w_i^{\pm} = \sum_j w_{ij}^{\pm}$, $w_j^{\pm} = \sum_i w_{ij}^{\pm}$, $w^{\pm} = \sum_{ij} w_{ij}^{\pm}$. Although it has been argued that $Q^+$ and $Q^-$ should be treated equally because positive and negative connections play different roles in functional brain networks (*45*), we defined an asymmetric formula of modularity as below based on previous studies which proposed that high-$Q^+$ modularity partitions are more optimal than high-$Q^-$ objectively (*45*),

$$Q = Q^+ + \frac{w^-}{w^+ + w^-}Q^-$$

$$= \frac{1}{w^+}\sum_{ij}\left[A_{ij}^+ - \gamma^+ \frac{w_i^+ w_j^+}{w^+}\right]\delta\left(g_i, g_j\right) - \frac{1}{w^+ + w^-}\sum_{ij}\left[A_{ij}^- - \gamma^- \frac{w_i^- w_j^-}{w^-}\right]\delta\left(g_i, g_j\right) \qquad (12).$$

The maximization of the modularity index $Q$ categorizes the nodes into the communities such that the total edge weight within the module is as large as possible. Hence, we selected one temporal window with the largest $Q$ over the several temporal windows. To measure changes in the modules during the learning, we calculated the *flexibility* of a node $f_i$ at each block. The *flexibility* was defined as the number of times the node changed modular assignment throughout the block, normalized by the number of all possible changes(*19*). The flexibility of the network in each block was calculated as below,

$$F = \frac{1}{N}\sum_{i=1}^{N} f_i \qquad (13),$$

where time window $N = 600/temporal\ scale$.

## Reference


1. T. E. Behrens, L. T. Hunt, M. W. Woolrich, M. F. Rushworth, Associative learning of social value. *Nature* **456**, 245-249 (2008).
2. T. E. Behrens, M. W. Woolrich, M. E. Walton, M. F. Rushworth, Learning the value of information in an uncertain world. *Nat Neurosci* **10**, 1214-1221 (2007).
3. E. C. Poulton, P. R. Freeman, Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin* **66**, 1-8 (1966).
4. D. H. Holding, Transfer between difficult and easy tasks. *British Journal of Psychology* **53**, 397-407 (1962).
5. E. Pulcu, M. Browning, The Misestimation of Uncertainty in Affective Disorders. *Trends Cogn Sci* **23**, 865-875 (2019).
6. E. Payzan-LeNestour, P. Bossaerts, Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput Biol* **7**, e1001048 (2011).
7. A. J. Yu, P. Dayan, Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681-692 (2005).
8. D. R. Bach, O. Hulme, W. D. Penny, R. J. Dolan, The known unknowns: neural representation of second-order uncertainty, and ambiguity. *J Neurosci* **31**, 4811-4820 (2011).
9. M. R. Nassar, J. T. McGuire, H. Ritz, J. W. Kable, Dissociable Forms of Uncertainty-Driven Representational Change Across the Human Brain. *J Neurosci* **39**, 1688-1698 (2019).
10. E. Payzan-LeNestour, S. Dunne, P. Bossaerts, J. P. O'Doherty, The neural representation of unexpected uncertainty during value-based decision making. *Neuron* **79**, 191-201 (2013).
11. M. Browning, T. E. Behrens, G. Jocham, J. X. O'Reilly, S. J. Bishop, Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* **18**, 590-596 (2015).
12. C. Gagne, O. Zika, P. Dayan, S. J. Bishop, Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *Elife* **9**, e61387 (2020).
13. J. T. McGuire, M. R. Nassar, J. I. Gold, J. W. Kable, Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* **84**, 870-881 (2014).
14. B. Massi, C. H. Donahue, D. Lee, Volatility Facilitates Value Updating in the Prefrontal Cortex. *Neuron* **99**, 598-608 (2018).
15. H. F. Clarke, G. J. Hill, T. W. Robbins, A. C. Roberts, Dopamine, but not serotonin, regulates reversal learning in the marmoset caudate nucleus. *J Neurosci* **31**, 4290-4297 (2011).
16. Z. M. Williams, E. N. Eskandar, Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat Neurosci* **9**, 562-568 (2006).
17. C. A. Seger, C. M. Cincotta, The roles of the caudate nucleus in human classification learning. *J Neurosci* **25**, 2941-2951 (2005).
18. C.-H. Kao *et al.*, Functional brain network reconfiguration during learning in a dynamic environment. *Nat Commun* **11**, 1682 (2020).
19. D. S. Bassett *et al.*, Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci U S A* **108**, 7641-7646 (2011).

20. D. S. Bassett, M. G. Mattar, A Network Neuroscience of Human Learning: Potential to Inform Quantitative Theories of Brain and Behavior. *Trends Cogn Sci* **21**, 250-264 (2017).
21. T. Xiang, T. Lohrenz, P. R. Montague, Computational substrates of norms and their violations during social exchange. *J Neurosci* **33**, 1099-1108 (2013).
22. G. B. Bissonette, M. R. Roesch, Development and function of the midbrain dopamine system: what we know and what we need to. *Genes Brain Behav* **15**, 62-73 (2016).
23. M. Haruno *et al.*, A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *J Neurosci* **24**, 1660-1665 (2004).
24. K. Samejima, Y. Ueda, K. Doya, M. Kimura, Representation of action-specific reward values in the striatum. *Science* **310**, 1337-1340 (2005).
25. S. M. McClure, G. S. Berns, P. R. Montague, Temporal Prediction Errors in a Passive Learning Task Activate Human Striatum. *Neuron* **38**, 339–346 (2003).
26. J. P. O'Doherty, P. Dayan, H. Critchley, K. Friston, R. Dolan, Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron* **28**, 329–337 (2003).
27. J. Jiang, J. Beck, K. Heller, T. Egner, An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. *Nat Commun* **6**, 8165 (2015).
28. R. Marsh, T. V. Maia, B. S. Peterson, Functional disturbances within frontostriatal circuits across multiple childhood psychopathologies. *Am J Psychiatry* **166**, 664-674 (2009).
29. K. Yuan *et al.*, Frontostriatal circuits, resting state functional connectivity and cognitive control in internet gaming disorder. *Addict Biol* **22**, 813-822 (2017).
30. C. Liston *et al.*, Frontostriatal microstructure modulates efficient recruitment of cognitive control. *Cereb Cortex* **16**, 553-560 (2006).
31. D. Meder *et al.*, Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. *Nat Commun* **8**, 1942 (2017).
32. U. Braun *et al.*, Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc Natl Acad Sci U S A* **112**, 11678-11683 (2015).
33. S. L. Bressler, V. Menon, Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn Sci* **14**, 277-290 (2010).
34. J. M. Shine *et al.*, The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron* **92**, 544-554 (2016).
35. E. A. Allen *et al.*, Tracking whole-brain connectivity dynamics in the resting state. *Cereb Cortex* **24**, 663-676 (2014).
36. J. Taghia *et al.*, Uncovering hidden brain state dynamics that regulate performance and decision-making during cognition. *Nat Commun* **9**, 2505 (2018).
37. A. N. Hampton, P. Bossaerts, J. P. O'Doherty, The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* **26**, 8360-8367 (2006).
38. C. D. Spielberger, R. L. Goruch, R. E. Lushene, P. R. Vagg, G. A. Jacobs, *Manual for the state-trait inventory STAI (Form Y1 – Y2)*. (Consulting Psychologists Press, Palo Alto, CA, 1983).

584 39. W. W. K. Zung, A Self-Rating Depression Scale. *Archives of General Psychiatry* **12**,
585 63-70 (1965).
586 40. J. H. Patton, M. S. Stanford, E. S. Barratt, Factor structure of the Barratt impulsiveness
587 scale. *Journal of clinical psychology* **51**, 768-774 (1995).
588 41. R. A. Rescorla, A. R. Wagner, *A theory of Pavlovian conditioning: Variations in the*
589 *effectiveness of reinforcement and nonreinforcement*. Classical conditioning II: Current
590 research and theory(BlackAH, Prokasy WF, eds) (Appleton-Century-Crofts, New
591 York, 1972).
592 42. M. E. Newman, M. Girvan, Finding and evaluating community structure in networks.
593 *Phys Rev E* **69**, 026113 (2004).
594 43. M. W. Cole, S. Pathak, W. Schneider, Identifying the brain's most globally connected
595 regions. *Neuroimage* **49**, 3132-3148 (2010).
596 44. M. E. Newman, Modularity and community structure in networks. *Proc Natl Acad Sci*
597 *U S A* **103**, 8577-8582 (2006).
598 45. M. Rubinov, O. Sporns, Weight-conserving characterization of complex functional
599 brain networks. *Neuroimage* **56**, 2068-2079 (2011).
600 46. V. A. Traag, J. Bruggeman, Community detection in networks with positive and
601 negative links. *Phys Rev E* **80**, 036115 (2009).

602

603

## Acknowledgements

**Funding:**

National Natural Science Foundation of China (31920103009)

National Natural Science Foundation of China (31871137)

Major Project of National Social Science Foundation (20&ZD153)

Young Elite Scientists Sponsorship Program by China Association for Science and Technology (YESS20180158),

Guangdong International Scientific Collaboration Project (2019A050510048)

Guangdong Key Basic Research Grant (2018B030332001)

Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (2019SHIBS0003)

Shenzhen Science and Technology Research Funding Program (JCYJ20180507183500566 and JCYJ20180305124819889)

## Author contributions

Conceptualization: P.X., H.A

Methodology: P.X., J.X

Investigation: J.X

Supervision: P.X., H.A

Writing—original draft: J.X., N.T.-D., Y.L., A.A., H.A., P.X

Writing—review & editing: J.X., N.T.-D., Y.L., A.A., H.A., P.X

## Declaration of interests

The authors declare no competing interests.

## Data availability

The data that support the findings of the current study are available from the corresponding author (P. X.) upon request.

## Code availability

The custom scripts used to analyze are available from the corresponding author (P. X.) upon request.

## Figures

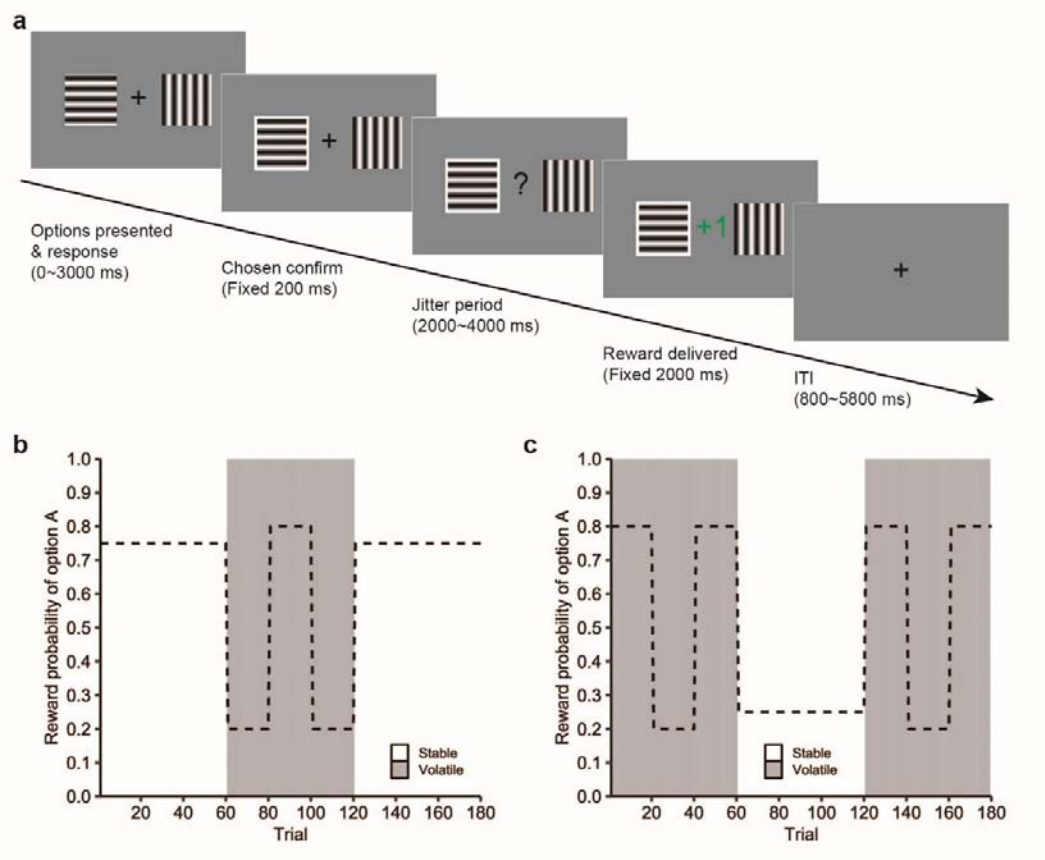


**Fig. 1. Task design. a) Experimental procedure.** Participants were required to choose one of the two options with either horizontal or vertical gratings to maximize reward. A response cue indicated which option had been selected. A variable jitter was followed by reward presentation. In the outcome phase, the reward (green "+ 1") or the no reward (red "+ 0") was presented for two seconds. **b) Reward probabilities across the course in the stable-volatile-stable task.** This task consisted of three blocks (stable-volatile-stable). In the stable block, one option was linked to a reward with 75% probability while the other one would be followed by a reward with 25% probability. In the volatile block, the reward probabilities of the two options would switch between 20% and 80% every 20 trials. **c) Reward probabilities across the course in the volatile-stable-volatile task.** This task consisted of three blocks (volatile-stable-volatile). The reward probability of options was same with the one in the stable-volatile-stable task.

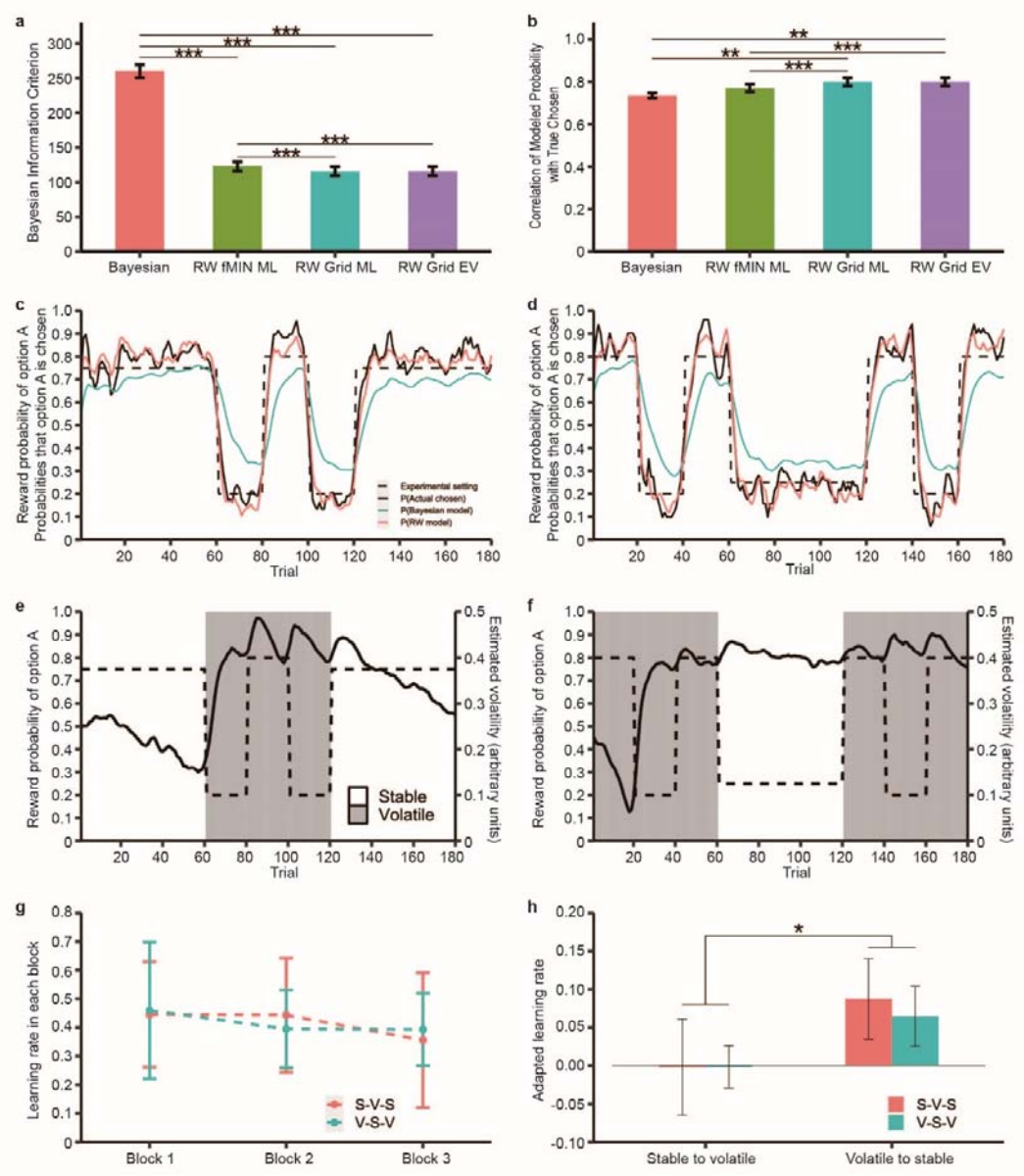


**Fig. 2. Behavioral results. a) Model comparison by Bayesian information criterion (BIC).** The RW Grid model shows the lowest BIC among all models. Error bars represent the standard error. **b) Model comparison by the correlation of modeled probability with participants' choices.** The RW Grid models showed the largest correlations among all models. Error bars represent the standard error. **c d) Learning curves illustrating participants' choices and estimates of model during the c) Stable-Volatile-Stable task** and **d) Volatile-Stable-Volatile task**. The black dashed line represents the experiment setting of the probability of the highly rewarded option. The black solid line represents the participant's choices. The salmon solid line represents the Rescorla-Wagner model prediction. The dark turquoise solid line represents the Bayesian model prediction. To illustrate participants' choices and model predictions, the variation across the participants and trials has been reduced by smoothing using a running average of four trials. **f) Environmental volatility estimated by the Bayesian model across the whole experiment. e) Stable-Volatile-Stable task**. In the stable-volatile-stable task, estimated volatility (black solid lines, righthand axes) decreased gradually in the stable states and increased

suddenly after each reversal. **f) Volatile-Stable-Volatile task**. In the volatile-stable-volatile task, after the first volatile state, estimated volatility (black solid lines, righthand axes) decreased slowly in the stable state. The underlying reward probabilities of the task are presented with black dashed lines (lefthand axes). **g h) Learning rates in each state. g) Learning rates fitted by the Rescorla-Wagner model to choices in each block of each task.** Salmon dashed line for stable-volatile-stable task. Dark turquoise dashed line for volatile-stable-volatile task. Dots represented the mean of participants' learning rates. Error bars represented the standard deviation of participants' learning rates in each block. **h) The adapted learning rates were significantly higher from volatile to stable state than it from stable to volatile state, regardless of the order in which the two blocks were first completed.** Error bars represent the standard error.

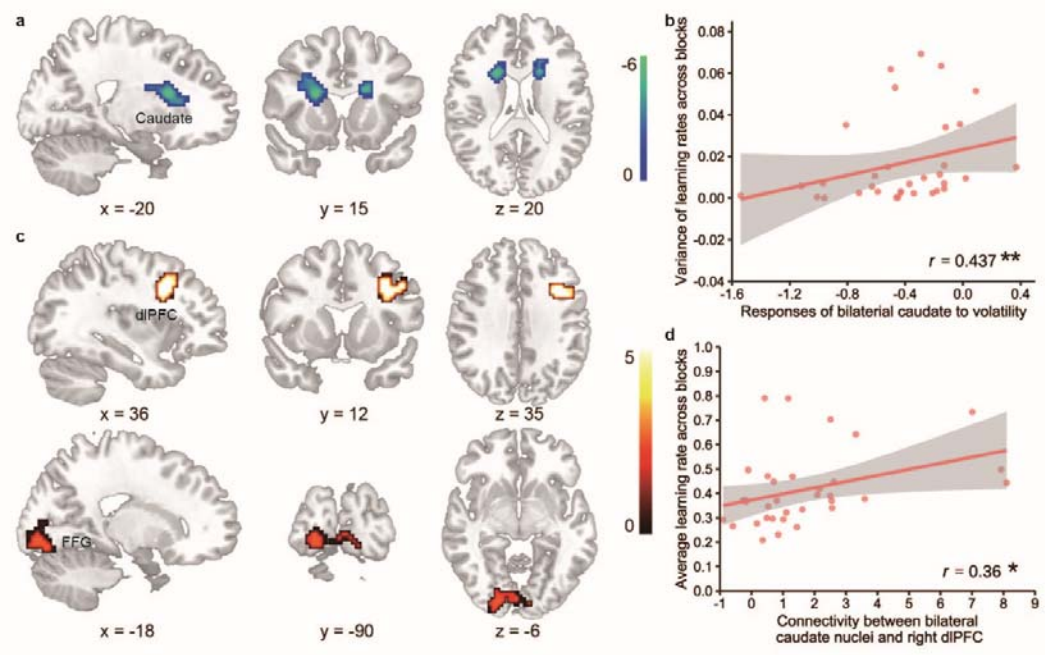


**Fig. 3. Brain activation. a) Responses of bilateral caudate nuclei to estimated volatility.** Activity of the bilateral caudate nuclei (x = -18, y = 12, z = 15) was negatively associated with the estimated volatility in outcome evaluation. **b) Volatility related brain activity predicated changes in learning rates across blocks.** The degree to which bilateral caudate nuclei tracked estimated volatility could predict the variance of the learning rates across the three blocks. **c) Brain regions showed significant functional connectivity with bilateral caudate nuclei modulated by environmental volatility. d) Correlation between caudate-related connectivity and learning rate.**

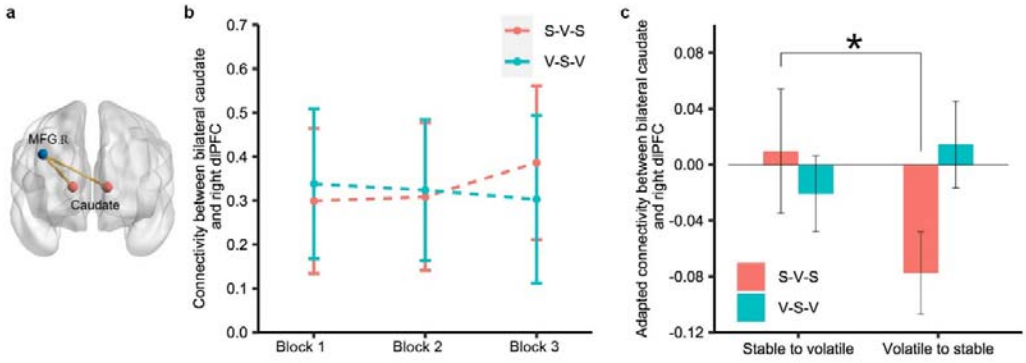


**Fig. 4. Dynamic functional connectivity. a) The locations of regions of interest (the caudate) and its target region (the MFG).** Both of the two regions were defined based on the Anatomical Automatic Labeling (AAL) atlas). **b) The connection between the caudate and middle frontal gyrus across blocks. c) Changes of connectivity between the caudate and dlPFC between transitions from volatile to stable and from stable to volatile environment.**

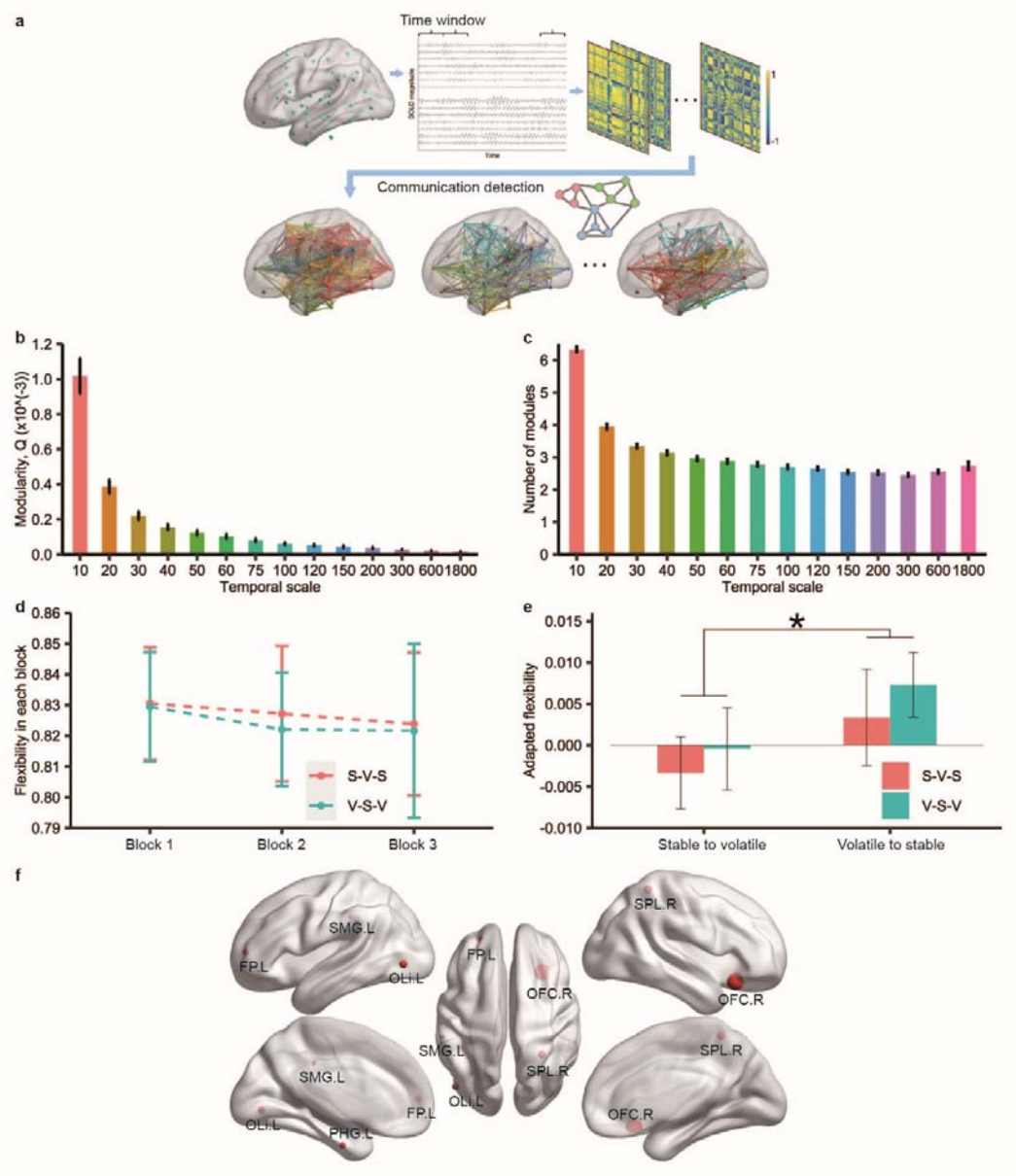


**Fig. 5. Dynamic brain networks. a) Schematic overview of the dynamic brain network analysis.** The modular architectures of functional connectivity were detected by the modularity index. **b) Modularity index *Q* of community across several temporal scales. c) The number of modules of community across several temporal scales.** Error bars represent the standard error. **d) Network flexibilities in three blocks during the learning were calculated at 10s temporal scale.** Salmon dashed line for the stable-volatile-stable task; Dark turquoise dashed line for the volatile-stable-volatile task. Dots represent the mean of flexibilities. Error bars represent the standard deviation of network flexibilities in each block. **e) The adapted flexibility between transitions from volatile to stable and from stable to volatile environments).** Error bars represent the standard error. **f) The brain regions represented adapted flexibility effect.** FP, frontal pole; SPL, superior parietal lobule; SMG, supramarginal gyrus; OLi, lateral occipital cortex; OFC, orbitofrontal cortex, PHG, parahippocampal gyrus; L, left; R, right.