The prefrontal–striatal signatures of reduced model-based learning in depressed patients

Xiaoxia Wang[1,2], Xiaoyan Zhou[3], Dong Zhang[4], Zongzhang Zhang[5], YuShun Gong[6], Zhengzhi Feng[7,*], Jing Ming Hou[8,*]

1. Department of Rehabilitation Medicine, Southwest Hospital, the first affiliated hospital to Army Medical University, Chongqing, China

2. Department of Basic Psychology, School of Psychology, Army Medical University, Chongqing, China

3. Department of Clinical Psychology, Mental Health Center of Chongqing, Chongqing, China

4. Department of Radiology, XinQiao Hospital, Army Medical University, Chongqing, China

5. National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

6. Department of Medical equipment and Metrology, College of Biomedical Engineering, Army Medical University, Chongqing, China

7. Department of Clinical Psychology, Southwest Hospital, the first affiliated hospital to Army Medical University, Army Medical University, Chongqing, China

8. Department of Rehabilitation Medicine, Southwest Hospital, the first affiliated hospital to Army Medical University, Chongqing, China

*Corresponding author: Zhengzhi Feng, E-mail: fzz@tmmu.edu.cn; Jingming Hou, E-mail: jingminghou@tmmu.edu.cn

Abstract

Background: Anhedonia is the core symptom of major depressive disorder (MDD). Accumulating evidence indicates that an imbalance between model-based (MB) and model-free (MF) reinforcement learning (RL) characterizes MDD, but the underlying neural substrates remain unclear. We examined whether alterations in MB and MF reward prediction error (RPE) neural signature underlie deficits in RL in depressed patients.

Methods: We used a two-stage Markov decision task (MDT) in combination with computational modeling to examine model-based and model-free learning. A total of 49 MDD and 41 matched HC individuals performed the MDT. 19 MDD and 21 HC individuals underwent functional neuroimaging during the MDT. The stress-RL deficits model was tested using a mediation model with MB and MF RL as mediators between stress and depressive/anhedonic symptoms.

Results: Depressed patients showed RL deficits, with less reliance on MB strategies and more reliance on MF strategies. MB and MF RL deficits mediated the relationship between stress and anhedonic symptoms, with specific striatal signatures (i.e., RPE_{MF} signals in VTA and caudate) mediating stress and anhedonia symptoms across MDD and HC groups.

Conclusions: This study showed deficits in model-based RL for depressed patients, with underlying neural deficits in prefrontal-striatal RPE signals, which would be promising for improving therapeutic practice in depression.

Keywords: Anhedonia; model-based learning; prefrontal-striatal circuits; functional neuroimaging; major depressive disorder

1 Introduction

Major depressive disorder (MDD) is one of the most prevalent mental illnesses and has high comorbidity and mortality worldwide [1, 2]. Anhedonia, defined as "decreased interest and pleasure in most activities most of the day" , is one of the cardinal symptoms of MDD according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5). Compared to other depressive symptoms, anhedonia is usually the last to be resolved [3]. However, the underlying mechanisms are not well understood. A more fine-grained conceptualization towards behavioral models as well as an understanding of the neural basis of anhedonia would improve the identification of specific targets of treatment [4].

Anhedonia could be conceptualized as the deficits in anticipating, experiencing and learning from reward with underlying neural substrates [5, 6]. Bodgan proposed reward processing as a putative intermediate phenotype of MDD [7]. The reward deficits model proposed that stressful events may disrupt the dopaminergic reward system and impair reward sensitivity, which is closely related to chronic stress-induced depression [8, 9]. Acute stress reduces reward responsiveness, especially for healthy individuals reporting greater anhedonic symptoms in their daily life [10], healthy individuals with genetic stress vulnerability [11] and MDD patients with high anhedonic symptoms [12], while perceived chronic stress predicts reduced reward responsiveness even when controlling for general distress and anxiety [13].

At the behavioral level, learning could be classified as habitual vs. goal-directed behaviors, depending either on reinforcement or cognitive representation of task. Reinforcement learning (RL) computational modeling with model-based (MB) and model-free (MF) dichotomies allows to capture these two components with finer perspective and bridge the gap between subjective experience and the neural substrates of anhedonia. The MF-RL is driven by prediction error (PE) while the MB-RL is driven by prediction error and mental simulation of task structure [14]. The MB-MF dichotomies extend beyond the operationalization of goal-directed and habitual behaviors, for

example, the spatial navigation [15]. The work with two-step decision-making task has revealed that humans combine MB and MF computations [16]. When faced with uncertain situations, individuals with depressive symptoms tend to adopt more MF and less MB learning strategies measured with other RL tasks [17, 18], similar with the results of humans [19, 20] and rats [21] under stress. Specifically, acute stress may shift individuals, especially those with higher levels of depression from the MB to MF strategy [22, 23]. Recent work emphasized the prediction error to stress could predict negative affective consequences after stress [24]. Increased MF strategies correlated with elevated depressive symptoms, and more MB strategies was linked to lower anhedonia levels across all patient groups (MDD, OCD and SZ) [25]. To summarize, the blunted ability in RL may result in deficiency in approach behavior, which is characteristic of anhedonia symptoms in depression [26].

At the neural level, dopamine encodes prediction error, which is signaled by phasic burst or functional neuroimaging [27]. Dopamine might enhance model-based control, whereas disruption of the prefrontal cortex could make behavior more habitual [28, 29]. More recently, neuroimaging studies have explicitly modeled RL with probabilistic reward tasks and suggested that MDD is characterized by impaired phasic reward prediction error (RPE) neural signals of reward circuits [30-32]. The blunted RPE responses to naturalistic stimuli was closely related to depression, who showed less emotional benefit from surprisingly good outcome (positive PE), in contrast to normal responses to surprisingly negative outcome (negative PE) [33], a phenomenon of positive emotional blunting regularly reported in depression. Furthermore, altered RPE signals have been correlated with increased anhedonia or depressive symptoms [31, 32]. The subclinical depression was associated with reduced PE in the insula (MB-RL) and caudate (MF-RL) [34]. Stress disrupted both MB (decreased hippocampal activity) and model-free (decreased lateral prefrontal activity) neural computation [35]. Additionally, the RPE signal in ACC could predict anhedonia symptoms one year later [36]. Anhedonia is correlated with resting-state graph theoretical metrics (e.g., Strength Centrality, Eigenvector Centrality and Local Efficiency) of RPE network (e.g., dACC, dlPFC, caudate and ventral striatum) functionally defined according to the Reward Flanker Task (RFT) [37]. To sum up, we speculate that alterations in RPE encoding in prefrontal-striatal circuits might underlie RL deficits in depression.

Therefore, the current study was conducted based on the following hypotheses: (1) depressed patients show a relative shift from model-based to model-free control; (2) RL (MB and/or MF) acts as a mediator between stress and depressive and/or anhedonic symptoms; and (3) impaired RL prefrontal-striatal neural signals (RPE) are correlated with depressive and/or anhedonic symptoms.

2 Materials and methods

2.1 Participants

All participants were recruited via posters in the community or advertisements in psychology classes. All participants provided written informed consent. This study received approval from the Ethical Committee of Army Medical University and complies with the ethical standards of the Helsinki Declaration.

(1) **Behavioral study**. 49 depressed patients and 41 healthy controls were recruited and included in the study to perform the behavioral task. Another subgroup of healthy young adults (n=183) was recruited for the replication of the results.

The inclusion criteria for both groups were as follows: (1) right handedness and (2) normal or corrected-to-normal vision aged 18-65 years old. Additional criteria for the depressed patients were the current experience of a depressed episode according to the DSM-5 criteria and mild to moderate depressive symptoms (Hamilton Depression Rating Scale (HAMD-17) score ≥17; Beck Depression Inventory (BDI) score ≥14). The depressed patient group was selected and diagnosed according to the structured clinical interview by psychiatrists in the Mental Health Center of Chongqing (CN) before the recruitment.

The exclusion criteria for both groups were as follows: (1) residual symptoms of or manifest axis-I disorders or (2) a history of learning disabilities, neurological illnesses or physical illnesses that significantly impair psychosocial functioning or brain function. Additional exclusion criteria for the healthy control group were a history of medication with antidepressants, antipsychotics, or tranquilizers or a history of any axis-I disorder.

(2) **Neuroimaging study**. Among the participants in study 1, 19 depressed patients and 21 healthy controls participated in study 2 to undergo the neuroimaging while performing the behavioral task.

2.2 Clinical measures

All recruited participants were administered clinical instruments in the interview before the experiment.

(1) **Beck Depression Inventory (BDI-II).** The BDI-II is a 21-item self-reported questionnaire measuring the severity of depression in normal and psychiatric populations. Each item is scored from 0 (not at all) to 3 (nearly every day) [38]. The translated Chinese version showed acceptable reliability (Cronbach's α coefficient=0.94, test-retest reliability coefficient=0.55) and validity (criterion validity with HAMD-17: r=0.67, P<0.01) [39]. To further separate depressive symptoms from anhedonic symptoms, the total score of the anhedonia items (4, 12, 19, 21) on the BDI was used as the anhedonic

symptom score (ASS), and the sum of the scores of the rest of the items was used as the depressive symptom score (DSS). These two scores were used additionally as variables to investigate the correlations between model-based and model-free RL signals and symptoms (*Section 3.5.2*).

(2) **Mood and Anxiety Symptom Questionnaire-Short Form (MASQ-SF**). The MASQ-SF is used to measure anxious and depressive symptoms [40]. It consists of 62 items which could be divided into 4 subscales. Each item is scored from 1 (not at all) to 5 (severe). The depression subscales include the general distress depression (GDD) and anhedonic depression (AD) subscales. The MASQ_AD subscale is used to assess trait anhedonia. The translated Chinese version of the MASQ-SF showed good reliability (Cronbach's α coefficient=0.94, test-retest reliability coefficient=0.82) and validity (four-factor model: NFI=0.90, CFI=0.90, GFI=0.92, RMSEA=0.07) [41].

(3) **Perceived Stress Scale (PSS)**. The PSS is used to assess the level of unexpected and unmanageable stress that an individual has experienced in the past month [42]. There are 10 items that are rated from 0 (never) to 4 (very common). The translated Chinese version has good reliability (Cronbach's α coefficient=0.91, test-retest reliability coefficient=0.69) and construct validity (two-factor model: NFI=0.91, CFI=0.97, GFI=0.93, RMSEA=0.04) in depressed, obsessive-compulsive and healthy individuals [43].

2.3 Behavioral study

We adopted the two-step Markov sequential decision task [16], which is designed to distinguish model-based and model-free RL. Before the formal task, all participants underwent a self-paced tutorial that introduced the task structure and provided 50 practice trials.

At the start of each trial, participants were shown a pair of abstract colorful stimuli (stage 1) and were asked to choose between them. The participant then reached stage 2 where they again had to choose between two abstract stimuli. The choice of one stage 1 stimulus more frequently led to one of two stage 2 stimulus pairs (P=0.70 or 0.30), while it rarely led to another stage 2 stimulus pair (P=0.30 or 0.70). The opposite was true for the other stage 1 stimulus. After choice of stage 2 choice participants received probabilistic feedback (reward in form of xxx currency) according to Gaussian random walks.

To ensure that our participants understood the instructions, we asked all participants about the transition structure ("Which red picture more frequently led to the two green pictures?"). There were 201 trials in total (**Figure 1**). After the task was completed, participants were paid extra bonuses upon their performance.

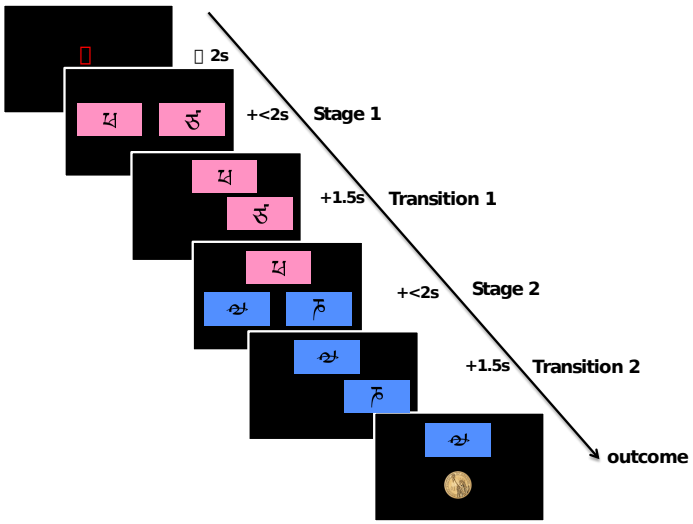


Figure 1 Flow chart of the two-stage Markov reward decision task

2.4 Neuroimaging study

After entering the 3.0T MRI scanning room, the participants lay supine on the scanning bed, with their head and body kept still. Rubber earplugs and headphones were used to reduce the impact of noise, and foam head pads were used to fix the head to reduce the head motion. T1-weighted structural images were collected from the whole brain with a FSPGR sequence using 3.0T General Electric (GE) Signa EXCITE scanner. The acquisition parameters were FOV= 25.6 cm, layer thickness 1.2 mm, resolution 1×1×1 mm, TE= 30 ms, turn angle =15°. T2-weighted functional images were collected by means of a gradient echo pulse sequence with 32 layers (sagittal), FOV= 19.2 cm, matrix size =64×64, spatial resolution =3×3×3 mm, flip angle =90°, TE=30 ms, TR=2000 ms.

2.5 Data analysis

The behavioral data were analyzed and computationally modeled as reflecting a combination of model-based and model-free reward learning using an adaptation of a previously described hybrid reinforcement learning (RL) model [16, 44]. The MB and MF components of RL were then entered into the mediation model to examine the indirect effects of stress on depressive/anhedonic symptoms with RL components as mediators. The impairments in neural substrates of prefrontal-striatal circuits underlying MB and MF RL of depressed patients were also examined.

2.5.1 Multi-level GLMM model of choice behavior

Generalized linear mixed effects model (GLMM) was used. The hierarchical logistic regression analysis was performed across the HC and MDD group, using the glmer function of lme3 package (Version 3.3.0.1959) in R software (Version 3.4.0). (1) The effects of transition and outcome on stay behavior for each group. The dependent variable (stay behavior) was a dichotomous

variable (1=stay vs. 0=shift). The independent predictors included transition (-0.5=rare vs. 0.5=common), and outcome (0.5=reward vs. -0.5=no reward). Subjects were taken as random effect. (2) The effects of transition and outcome on stay behavior across groups. According to previous studies [44, 45], the dependent variable (stay behavior) was a dichotomous variable (1=stay vs. 0=shift). The independent predictors included group (HC vs. MDD), transition (-0.5=rare vs. 0.5=common), and outcome (0.5=reward vs. -0.5=no reward). Group (HC vs. MDD) was taken as fixed effect. Within-subject factors (intercept, the transition, outcome, and their interactions) were taken as random effects. The main effect of outcome (i.e., the model-free term) and transition-by-outcome interaction (i.e., the model-based term) were the effects of interest. The intercept reflects the tendencies to repeat the action from the previous trial [46], using bound-constrained Optimization by Quadratic Approximation (bobyqa) with 1e6 functional evaluations. The model was defined as follows: Stay ~ 1 + transition × outcome × Group + (1 + transition × outcome|Subject). The trials in which a participant missed the option during the first- or second-stage were removed.

2.5.2 Individual-level RL computational model of choice behavior

The modelling was performed with Matlab 2019b (https://www.mathworks.com) , with the emfit toolbox developed by Huys et al. [47] within and across the groups, i.e., MDD, healthy control (HC) and young healthy adult (YHA) groups. The task consisted of three states: first stage S_A; second stage S_B and S_C, each resulting from two actions (a_A and a_B). Then the transition between conditions were as follows: (1) common transitions. $P(S_B|S_A,a_A) = 0.7$, $P(S_C|S_A,a_B) = 0.7$, or vice versa, (2) rare transitions. $P(S_B|S_A,a_A) = 0.3$, $P(S_C|S_A,a_B) = 0.3$. The goal of RL modeling is to map each state-action pair to its expected future value. The model-free value was modeled with the State-Action-Reward-State-Action (SARSA) algorithm. This algorithm updates the action value for each state-action pair according to:

$$Q_{MF}(S_{i,t},a_{i,t}) = Q_{MF}(S_{i,t},a_{i,t}) + \alpha_i\delta_{i,t} \quad (1)$$

where $\delta_{i,t} = r_{i,t} + Q_{MF}(S_{i+1,t},a_{i+1,t}) - Q_{MF}(S_{i,t},a_{i,t})$ (2)

The model-based value was modeled by learning a transition function and immediate reward values for each state and computing cumulative state-action values by iterative prediction of the reward values.

$$Q_{MB}(S_A,a_j) = (S_B|S_A,a_j) \max_{a \in \{a_A,a_B\}} Q_{MF}(S_B,a) + P(S_C|S_A,a_j) \max_{a \in \{a_A,a_B\}} Q_{MF}(S_C,a) \quad (3)$$

The weighted sum of the model-free and model-based values was calculated

for the first stage:

$$Q_{net}(S_A,a_j) = \omega Q_{MB}(S_A,a_j) + (1-\omega)Q_{MF}(S_A,a_j) \quad (4)$$

where w is the weighting parameter, which was assumed to be constant across trials.

In the first and second stages, the probability of action can be represented as the net state-action value Q_{net}, inverse temperature parameter β_1 and β_2 (which differ between stage 1 and 2), perseverance parameter p and indicator function rep(a):

$$P(a_{i,t} = a|s_{i,t}) = \frac{\exp(\beta[Q_{net}(s_{i,t},a) + p \cdot rep(a)])}{\sum_{a'} \exp(\beta_i[Q_{net}(s_{i,t},a') + p \cdot rep(a')])} \quad (5)$$

The update of the first-stage action value by the second-stage prediction error at the end of each trial according to eligibility trace parameter λ, $Q_{MF}(S_{1,t},a_{1,t}) = Q_{MF}(S_{1,t},a_{1,t}) + \alpha_1\lambda\delta_{1,t}$. In the first and second stages, different learning rates α_1 and α_2 were adopted to allow for potential differences in learning from state transitions vs. rewards [16].

The estimation was derived from simulations of SARSA and model-based algorithms using the parameters with the best fit to the participants' data. We estimated the seven free parameters (α_1, α_2, β_1, β_2, λ, ω, p) separately for each subject to maximize the negative log likelihood (LL) of the obtained choices given the previously observed choices and rewards summed over all participants and trials. The model fitting procedures were performed by Expectation-Maximisation (EM) to find group priors and individual (Laplace) approximate posterior distributions for the estimates for each parameter for each participant.

Bayesian model comparison at the group level is to assess the model parsimony by comparing posterior probability of each model given the dataset for all participants. The individual parameters were integrated out by sampling from the fitted priors and computing the posterior log likelihood (LL) of each model given all the data, with Bayesian information criterion (BIC) at the group level (with smaller BIC representing greater model parsimony). We first compared all computational parameters between groups (MDD vs. HC) based on the default hybrid model (7 parameters, MB and MF), and then tested whether the MDD and HC groups showed differences in their best fitting model.

2.5.3 Stress-RL deficits mediation model of depression

The mediating role of RL deficits in the relationship between stress and depression was analyzed in depressed patients. Using SPSS 19.0 (https://www.ibm.com/cn-zh/spss) combined with the PROCESS macro (model 4) [48], stress was taken as the independent variable, depressive symptoms as the dependent variable, model-based and model-free RL were used as the

mediating variables to investigate whether stress affected depression and/or anhedonic symptoms through the effects of model-based and/or model-free RL.

2.5.4 The prefrontal-striatal RL BOLD signals

SPM12 (Wellcome Trust Center for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/) was used for magnetic resonance imaging data preprocessing and analysis.

Preprocessing. ① Slice timing. The scanning sequence was interlayer scanned (1:2:31, 2:2:32), including 32 layers with the reference layer of 16. TR=2, TA=2-2/32. ② Realignment. Head motion correction was performed using least squares method and space transformation with 6 parameters (rigid body model), and the allowable head motion range (translation ≤3.0mm, rotation ≤3.0°). ③ Normalization. The functional images were normalized to the Montreal Neurological Institute (MNI) template, with resampling resolution of 2×2×2 mm^3, bounding box [-78 112-70; 78 76 85], and other default parameters. ④ Smoothing. The Gaussian kernel of 8 mm Full width at half maximum (FWHM) was selected for smoothing.

Task-related functional image analysis. The general linear model (GLM) was used for ROI-based analysis, the first 7 volumes of each scan were removed, and the head motion parameters were included as covariates. *First*, the RPE was calculated at the beginning of the second stage ($\delta_{1,t}$) and at the time of the reward outcome ($\delta_{2,t}$), with their general form expressed as equation (2). *Second*, the RPE was calculated as the partial derivative with respect to ω, which stands for the difference regressor between the RPEs with respect to model-based and model-free actions. The nuisance regressors included $P(a_{1,t}|s_A)$ from equation (5) and its partial derivative with respect to ω [16].

Finally, a standard hemodynamic response function (HRF) was constructed to locate the RPE BOLD signals of prefrontal-striatal ROIs within and across groups (HC vs. MDD). The areas of interest (ROIs) were defined using PickAtlas v3.0.5 (https://www.nitrc.org/projects/wfu_pickatlas). The ROIs include: (1) medial (including BA25, 12) and lateral (including BA10, 11, 47) orbitofrontal lobes (i.e., MOFC, LOFC)(IBASPM71); and (2) midbrain/VTA (TD lobes); (3) ventral striatum/nucleus accumbens (i.e., VS/NA) (IBASPM71); (4) dorsal lateral and medial striatum (i.e., putamen, caudate).

The second level analyses for RPE BOLD signals of ROIs within and across groups were conducted**.** (1) One-sample t test. The model-based and model-free learning weight ω was used as covariate, and the significant RPE_{MB} and RPE_{MF} signals within the prefrontal-striatal ROIs in HC and MDD group respectively were examined. (2) Two-sample t test. The differences in RPE_{MB} and/or RPE_{MF} signals between HC and MDD groups were examined.

Correlation analyzes. Correlations between RPE ROI signals and symptoms (depression, anhedonia) were explored. The Pearson correlations between RPE_{MB} and/or RPE_{MF} ROI signals and depressive (BDI and DSS) symptoms across groups were examined. The significant correlation coefficients ($P<0.05$) were reported.

Mediation model. We hypothesized that reward prediction error (RPE) brain signals mediate the relationship between model-based and/or model-free RL and severity of anhedonia. RL was the independent variable, severity of anhedonia the dependent variable, RPE brain signals of ROIs (including lateral and medial OFC, putamen, caudate, VTA and NAc) the mediators. The PROCESS macro in SPSS 19.0 (model 4) (https://www.ibm.com/cn-zh/spss) was utilized to examine the mediating role of RPE BOLD signal of ROIs between RL behavior and anhedonia. A bootstrapping analysis was conducted with 1000 resamples to calculate the bias-corrected 95% confidence intervals (CI) and test the significance of the mediation effect. The 95% CIs that did not include zero were regarded as significant [48].

The significance level is set at 0.05, and the cluster size is set at ≥5. The brain function activation map based on the peak voxel coordinates and significant clusters was reported, using xjview96 (http://www.alivelearn.net/xjview/) toolbox and the small volume correction (SVC) was adopted (P<0.05, extent threshold k=5 voxels), restricting the correction to the prefrontal ROIs (including lateral and medial OFC, putamen, caudate, VTA and NAc). For display purposes in the supplementary materials, we did whole-brain analyses and rendered activations at an uncorrected threshold of p<.0001, in combination with a cluster-based family wise error (FWE) correction of $P_{FWE}<0.05$ (i.e.,>336 voxels). MRIcro was used to output brain function activation map which was superimposed on the standard brain structure map (ch2bet).

3 Results

The individual- and multi-level behavioral analyses of choice behavior confirmed reduced model-based reward learning in depressed patients. The MB and MF components of RL were then entered into the mediation model to examine the indirect effects of stress on depressive/anhedonic symptoms with RL components as mediators. The impairments in neural substrates of prefrontal-striatal circuits underlying MB and MF RL of depressed patients were also examined.

3.1 Demographic and clinical data

There was no significant difference between the HC and MDD group (study 1) in the gender ratio ($\chi^2=1.07$, $df=2$, $P=0.30$) or education level ($\chi^2=6.48$, $df=3$, $P=0.09$). There were statistically significant differences in depression (BDI) (t=-7.44, $P<0.001$) and anhedonia (MASQ_AD) (t=-7.49, $P<0.001$) scores

between the HC and MDD group. There was a statistically significant difference in stress (PSS) scores between the HC and MDD group (t=-5.21, P<0.001) (**Table**).

Table 1 Descriptive statistics of the demographic and clinical information

	Healthy controls (HC, n=41)	Depressed patients (MDD, n=49)	Young healthy adults (YHA, n=183)
Demographics			
age (M±SD)	38.24±7.62	40.10±10.58	22.36±2.88
gender ratio (M/F)	14/27	22/27	159/24
education (M±SD)	2.91±0.83	2.38±0.95	4.04±0.21
Symptoms			
BDI (M±SD)	7.66±10.09	25.79±13.60	5.74±8.12
MASQ_AD (M±SD)	49.20±15.31	72.61±13.69	48.87±15.00
RSAS (M±SD)	8.30±5.98	16.24±8.87	7.34±5.27
Stress			
PSS (M±SD)	23.80±7.63	31.13±5.10	22.96±6.28

Note: Education level is an ordinal variable that is divided into four categories: junior high school (1), senior high school (2), college degree (3) and postgraduate degree (4).

3.2 The effect of depression on model-based and model-free RL

The stay probabilities for common vs. rare transition, with reward vs. nonreward outcome across groups were examined (**Figure 2**).

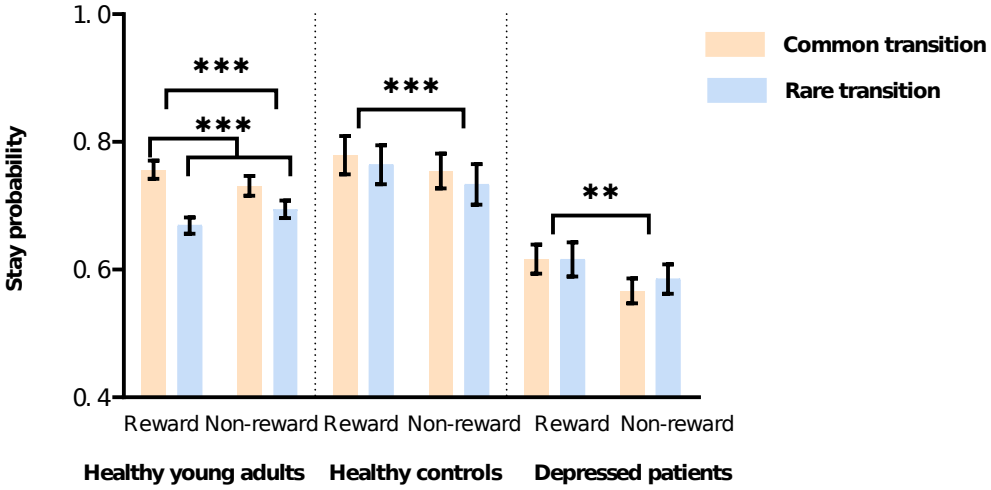


Figure 2 Stay probability of different groups (HYA, HC and MDD) (Mean±SEM)

Across the groups, the logistic regression analyses yield that: (1) the main effect of the outcome was statistically significant in the HC group (β=0.14, *P*<0.001) and the depressed patients (β=0.11, *P*=0.007); (2) the main effect of the outcome was statistically significant (β=0.25, *P*<0.001), and the transition-by-outcome interaction was statistically significant (β=0.08, *P*<0.001) in the HYA group. The results suggested that the HC and MDD group mainly adopt model-free RL, while the HYA group adopt both model-free and model-based RL. Across the three groups, the intercept item had a statistically significant main effect (*P*<0.001), suggesting that all groups had a tendency to repeat the first stage selection of the previous trial, which was independent of reward outcome or transition condition (**Table 2**).

Table 2 Generalized linear mixed-effects model of reward outcome, transition type and anhedonic symptoms on stay behavior across groups

Predictor	**Estimates of effects (SE)**	Z	***p***	**Model evidence (AIC, BIC, -LL)**
HC				
Intercept	1.54(0.20)	7.59	<0.001***	7446.8, 7544.5, -3709.4
Outcome	0.14(0.06)	2.43	0.02*	
Transition	0.07(0.04)	1.67	0.09	

Outcome×Transition	0.04(0.04)	0.96	0.34	
HYA				
Intercept	1.18(0.08)	13.98	<0.001***	
Outcome	0.25(0.03)	9.54	<0.001***	36855.0,
Transition	-0.002(0.02)	-0.13	0.89	36973.5, -18413.5
Outcome×Transition	0.08(0.02)	4.70	<0.001***	
MDD				
Intercept	0.44(0.10)	4.46	<0.001***	
Outcome	0.11(0.04)	2.70	0.007**	9590.9,
Transition	-0.008(0.03)	-0.27	0.78	9687.5, -4781.5
Outcome×Transition	0.02(0.03)	0.67	0.50	

Note: *P<0.05; **P<0.01; ***P<0.001. SE=standard error; AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion; -LL=negative log-likelihood.

Notably, although the main effect of transition within each group was not statistically significant, the main effects of participants' group, outcome as well as the group-by-transition interaction were statistically significant across groups (MDD and HC) (Ps≤0.05) (**Table 3**). The rewarded trials tend to induce more stay behaviors across groups, which was similar with within-group results. Furthermore, the main effect of group suggested that depressed patients are less inclined to choose the stay behavior in the first stage, and this tendency persists even when the previous trial is common transition or rewarded.

Table 3 Generalized linear mixed-effects model of reward outcome, transition type and group on stay behavior

Predictor	**Estimates of effects (SE)**	Z	***p***	**Model evidence (AIC, BIC, -LL)**
Intercept	1.54(0.20)	7.56	<0.001***	18176.5,

Outcome	0.15(0.06)	2.43	0.015*	18515.0, -9044.3
Transition	0.07(0.04)	1.67	0.09	
Outcome×Transition	0.04(0.04)	0.96	0.337	
Group	-1.08(0.22)	-4.75	<0.001***	
Group×Outcome	-0.03(0.07)	-0.41	0.68	
Group×Transition	-0.10(0.05)	-1.99	0.046*	
Group×Outcome×Transition	-0.03(0.05)	-0.67	0.50	

Note: *P<0.05; **P<0.01; ***P<0.001. SE=standard error; AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion; -LL=negative log-likelihood.

3.3 Computational RL Model fit within and between groups

3.3.1 The default RL model parameters between groups

As default, we first adopted the assumption that the hybrid model is the best fitting model across all participants (which was only supported in the HC group according to Section 3.3.2, Figure 2). The Shapiro-Wilk test was used to investigate whether the RL parameters of each group were normally distributed. Those participants with too much missing data were deleted from the analyses (n=5). For the parameters estimated from the behavioral data, values more than two standard deviations (95% confidence intervals) above the group mean value are marked as outliers and also excluded from the analyses.

The results showed that parameters were not normally distributed (P<0.05). As a result, the nonparametric Mann-Whitney U test was used to investigate whether there were significant intergroup differences in the RL parameters.

For simplicity and in convenience of comparison, we listed the RL parameters for both groups as the default hybrid model (MB and MF combined, 7 parameters). The depressed patients had a significantly lower learning rate (α_2), lower model-based learning weights (ω) and more frequent transitions of options (ρ) than the healthy control group. The model evidence (Bayesian information criterion, Laplace approximation of Bayes factor, and negative log-likelihoods) of the behavioral data fitted by the seven-parameter hybrid model between the two groups were statistically significant (Ps<0.001), with better model fit of the hybrid model for HC and YHC group (**Table 4**).

Table 4 Hybrid reinforcement learning model parameters for each group

Parameter	MDD (n=44)			HC (n=41)			Z (MDD vs. HC)	*P*	YHC (n=176)			Z (MDD vs. YHC)	*P*
	25%	50%	75%	25%	50%	75%			25%	50%	75%		
α1	1.698	2.301	2.509	1.706	1.913	2.294	-1.504	0.133	1.577	1.930	2.371	-0.175	0.861
α2	0.726	1.151	1.678	-0.261	0.692	1.456	-2.436	0.015*	0.746	1.195	1.590	-3.352	0.001**
β1	-3.032	-2.307	-0.960	-3.809	-2.822	-1.870	-2.331	0.02*	-2.612	-1.778	-0.271	-4.380	<0.001**
β2	-3.110	-1.806	-0.162	-3.745	-2.384	-0.167	-1.222	0.222	-2.761	-1.007	0.102	-2.351	0.019
λ	-0.784	0.077	0.634	-1.839	-1.147	-0.449	-3.272	0.001**	-1.038	0.047	0.946	-4.337	<0.001**
ω	-1.976	-1.301	-0.667	-2.593	-1.806	-1.093	-2.964	0.003**	-1.625	-1.137	-0.536	-4.396	<0.001**
ρ	0.015	0.057	0.120	-0.002	0.016	0.043	-3.966	<0.001**	0.025	0.083	0.170	-5.672	<0.001**
BIC	494.831	451.153	368.533	450.465	349.374	265.124	-3.412	0.001**	461.231	390.547	298.990	3.923	<0.001**
Lap	269.954	250.702	210.660	252.125	202.020	159.735	-3.085	0.002**	222.453	180.786	145.637	3.835	<0.001**
-LL	202.828	244.138	265.977	151.124	193.249	243.794	-3.412	0.001**	168.057	213.835	249.177	3.923	<0.001**

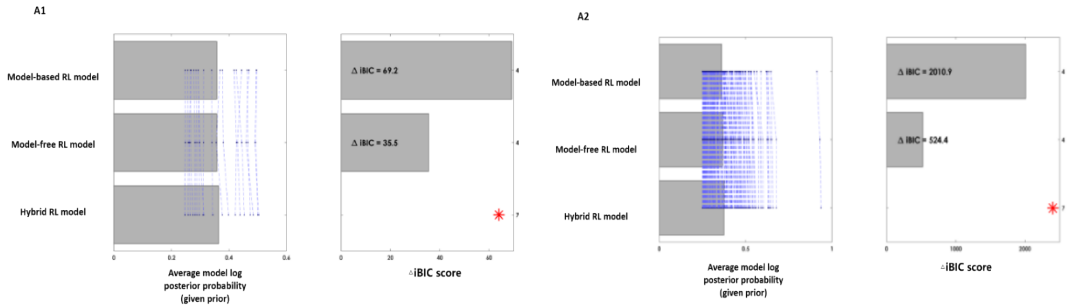
Note 1: α1, α2, learning rates of stage 1 and stage 2, respectively; β1, β2, choice repetition of stage 1 and stage 2, respectively; λ, reinforcement eligibility trace; ρ, perseverance parameter; ω, model-based versus model-free parameter weights; BIC, Bayesian information criterion; Lap, Laplace approximation of Bayes factor; -LL, negative posterior log-likelihood estimation. Multiple comparison was corrected by the Bonferroni method, and the significance threshold was set at $P_{Bonferroni}$s<0.007. *** P<0.001, ** P<0.01, * P<0.05

Specifically, the parameter (λ) of depressive patients is significantly lower than that of the normal control group, and is close to 0 (**Table 2**). The parameter (λ) can reflect the relative influence of the action value of the second stage and the final reward outcome on the choice of the first stage. If λ=1, it means that only the final reward outcome affects the choice of the first stage (model-free learning). If λ=0, it means that only the action value of the second stage affects the choice of the first stage (model-based learning). The above results suggest that the final reward outcome exerts less influence on the first stage action selection of depressed patients, showing deficits in MF RL.

Depression patients' model-based RL weight (ω) decreased (**Table 2**), mainly adopted MF learning strategies, and depend less on changes in environmental state (e.g., transition probability) to modify the action strategy. Additionally, the second stage learning rate (α) decreased (**Table 2**), which revealed that the degree of reward prediction value updates slower in depressed patients, which then may result in decreased stay behavior for the next trial.

3.3.2 The RL model comparison within groups

Using random-effects Bayesian model selection (BMS) [49] for different groups, the results showed that the hybrid RL model (MB and MF) best fit the behavioral data of the HC group (inside-scanner) (**Figure 3, A1**), while the MF RL model best fit the behavioral data of the depressed patients (inside-scanner) (**Figure 3, B1**). When all healthy participants, HC (n=219) (inside & outside-scanner) were included, similar results were found (**Figure 3, A2**), suggesting the hybrid RL model fit across healthy individuals. To replicate the results, we also conducted model comparison for all depressed patients (n=43, from study 2), which yielded similar results of best MF model fit (**Figure 3, B2**).



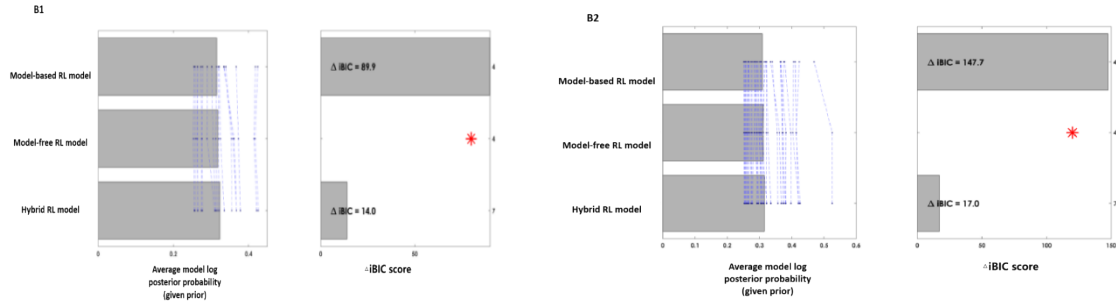


Figure 3 Comparison of the degree of model fit of different computational models: the models with the optimal fit for the healthy control group and depressed group were the mixed model and the model-free RL model, respectively. Note: A1: model evidence for the HC group (n=19, inside-scanner) (-LL, △iBIC); B1: model evidence for the MDD patients (n=17, inside-scanner) (-LL, △iBIC); A2: model evidence for all healthy participants (HYA and HC) (n=219, inside & outside-scanner) (-LL, △iBIC); B2: model evidence for all MDD patients (n=43, inside & outside-scanner) (-LL, △iBIC). The dotted lines represent the estimated parameters of each subject.

Collectively, the above results were consistent with the model evidence and intergroup comparisons based on RL weights (**Table 2**). The model evidence suggested that for the hybrid model used to fit the behavioral data of the participants, the degrees of model fit (BIC, LAP and -LL parameters) of the HC group were better than those of the MDD group.

3.4 The mediating role of RL between stress and anhedonia/depression

The results showed that for depressed patients, stress could influence anhedonic symptoms (MASQ_AD) via model-based (β=-37.73, SE=16.64, *P*=0.03, 95% CI=-71.38~-4.08) and model-free (β=-17.73, SE=16.64, *P*=0.03, 95% CI=-71.38~-4.08) RL. However, no significant mediating role was found for either model-based (β=-139.50, SE=96.25, *P*=0.16, 95%CI=-334.18~55.18) or model-free (β=-20.75, SE=17.43, *P*=0.24, 95%CI=-56.01~14.52) RL between stress and depressive symptoms.

3.5 The prefrontal-striatal neural substrates of RL in depression

For the MDD group, the RPE_{MF} signals in LPFC, midbrain/VTA and LOFC were less activated than those in the HC group. The RPE_{MF} signal in LPFC was greater than in the HC group (**Table 5, Figure 4**). No significant activation with RPE_{MF} was found for the between-group comparison of whole-brain analyses (cluster based P_{FWE}<0.05).

Table 5 Model-free RPE in PFC-striatal circuits between groups

Region	Side	MNI Coordinates			Cluster size	t score
		x	y	z		

HC>MDD						
LPFC	L	-52	20	12	33	2.733
Midbrain/VTA		0	-20	-14	25	2.196
LOFC	L	-32	26	-20	7	1.850
MDD>HC						
LPFC	R	44	8	28	10	2.036
LPFC	L	-46	8	26	5	1.957

Note:ROI-based analyses, P<0.05, extent threhold k=5 voxels, small volume corrected.

①HC>MDD

(A) Left LPFC **(B) Midbrain** **(C) Left LOFC**

②MDD>HC

(C) Right LPFC **(D) Left LPFC**

Figure 4 The differences of Model-free PRE in PFC-striatal circuits between groups

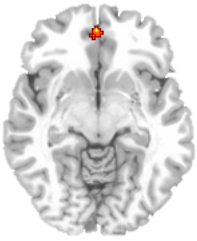
In addition, for the MDD group, the RPE_{MB} signals in the LPFC, MPFC, LOFC, MOFC, midbrain, and dorsal striatum (putamen, caudate) were less activated than those in the HC group (**Table 6, Figure 5**). The RPE_{MB} signals in the LPFC and MPFC were more activated than those in the HC group (**Table 6, Figure 6**). No significant activation with RPE_{MB} was found for the between-group comparison of whole-brain analyses (cluster based P_{FWE}<0.05).

Table 6 Model-based RPE in PFC-striatal circuits between groups

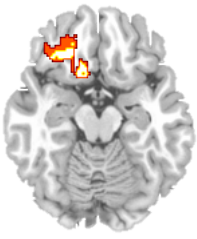
Region	Side	MNI Coordinates			Cluster size	t score
		x	y	z		
HC>MDD						
LPFC	R	42	38	34	232	3.387
MPFC	L	-2	48	-10	31	2.249
LOFC	L	-30	30	-18	278	3.875
LOFC	R	42	30	-4	15	2.278
MOFC	L	-8	20	-18	25	2.594
MOFC	L	-2	52	-10	5	1.949
Midbrain	R	4	-14	-4	76	2.664
Putamen	L	-14	8	-6	11	2.026
Caudate	L	-12	10	-6	13	2.131
MDD>HC						
LPFC	L	-44	42	10	45	2.224
LPFC	R	44	22	42	5	1.886
MPFC	R	2	30	60	45	2.598
MPFC	L	-6	46	46	17	1.984

Note:ROI-based analyses, $P<0.05$, extent threhold k=5 voxels, small volume corrected.

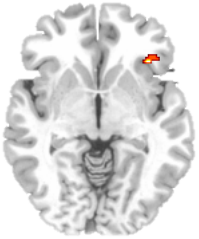
(A) Right LPFC



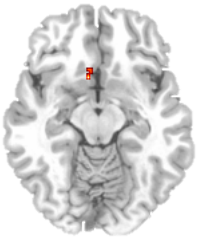
(B) Left MPFC



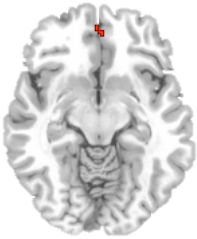
(C) Left LOFC



(D) Right LOFC



(E) Left MOFC



(F) Left MOFC

(G) Midbrain **(H) Putamen** **(I) Caudate**

Figure 5 The differences of model-based PRE in PFC-striatal circuits between groups (HC>MDD)

(A) Left LPFC **(B) Right LPFC**

(C) Right MPFC **(D) Left MOFC**

Figure 6 The differences of model-based PRE in PFC-striatal circuits between groups (MDD>HC)

4 Discussion

The current study examined the effects of depression/anhedonia on model-based and model-free RL. Our results suggested that depressed patients were inclined to search for other possible rather than stay with their original strategies. They showed a relative shift from model-based to model-free control. The decreased activation in prefrontal-striatal circuits (L/MOFC, VTA and DS) was shown in the depressed patients during model-based and model-free RL.

4.1 The reduced MB learning in depression

The results revealed that the reward outcome rather than the outcome-by-transition interaction significantly influenced the first-stage

choice of depressed patients, suggesting that depressed patients mainly adopted model-free RL strategies, with nonsignificant differences from the healthy controls. These results suggested that the model-free RL of depressed patients was relatively unimpaired. Computational modeling results showed that the hybrid model (MB and MF) best fit the choice behavior of the healthy control group, while the MF model best fit the choice behavior of depressed patients, which supported hypothesis (1). This study also showed that depressed patients' propensity to switch from previous behavior choices persisted even when participants were faced with positive reinforcement or the presence of common transitions in the environment (more frequent transitions of options (ρ) than the healthy control group). This is consistent with previous findings that when faced with an uncertain environment, individuals with depressive symptoms tend to explore alternative options in a disorderly manner rather than sticking to original action strategies to maximize expected benefits [18]. Presumably, this may reflect the decreased representation of task structure in depressed individuals. Depressed patients may reduce the complexity of state space sampling by reducing the number of differentiable states, shifting the balance between model-free and model-based strategies toward the former [50]. This assumption was supported by the neuroimaging results of depressed patients in which RPE_{MB} signals of OFC (encoding task state space, see Section 4.4.2) [51] were less activated when compared to HC group.

4.2 Stress-RL deficits model of anhedonia

The study found that MB and MF RL played the mediating role between stress and anhedonic symptoms, which confirmed the hypothesis (2) and consistent with previous evidence that physiological and psychosocial stress may affect individual MB RL but not MF RL [20, 22]. In particular, chronic stress combined with acute psychosocial stress reduces MB behavioral control, while acute psychosocial stress alone does not change MB behavioral control [20]. Recent evidence showed that chronic stress may decrease the model-free behaviors (lever press) as well as model-based behaviors (high fixed-rate reinforcement identification) during instrumental conditioning of rats [52]. Previous evidence suggests that stress affects the structure and function of the prefrontal cortex [53], which may result in the shift from model-based to model-free control [19, 54]. Furthermore, model-based RL deficits result in a reduced sense of controllability, with stress-induced serotonin responses and changes in adaptive behavior (such as reduced approach coping and increased avoidance coping), leading to a decrease in contact with positive reinforcement, which may then exacerbate anhedonia [55, 56]. Above all, these studies suggest that the cumulative effect of chronic stress may be a risk factor leading to MB and MF RL deficits.

4.3 The neural deficits in RL for depression

Model-based brain function imaging showed that in the healthy control group, there were RPE_{MF} signals in NAc, putamen, L/MOFC, and RPE_{MB} signals in LPFC, consistent with previous findings [57-59]. The comparison between groups revealed deficiency in the prefrontal (OFC, MPFC) and striatal (VTA, DS, NAc) regions as well as aberrant functional connectivities underlying model-free and model-based RL, which is consistent with our hypothesis (3).

4.3.1 The neural deficits in model-free RL for depression

VTA↓ and VTA-NAc↑. *First*, this study found reduced RPE_{MF} signal of VTA in the depressed patients,which is consistent with previous studies [31, 60]. Additionally, we did not find altered RPE_{MF} signal of NAc in depressed patients, consistent with previous study [61]. We also found that the RPE signal in NAc is negatively correlated with anhedonia, consistent with previous studies [31]. NAc is involved in transformation of reward value to action [62]. However, for depressed patients, the signal could not be used to update the reward expectancy, which is critical for the transformation into action value. Moreover, this deficiency worsened with increased anhedonia in depressed patients [61]. *Second*, as a complement, our study found statistically significant positive VTA-NAc FC in the MDD group (r=0.941, P<0.001), as compared to no statistically significant FC in the HC group (r=0.128, P=0.600). VTA projects dopamine-releasing neurons to target areas such as NAc (the mesolimbic pathway) and MPFC (the mesocortical pathway). Therefore, reduced VTA activation may lead to compensatory increase of VTA-NAc functional connectivity in depressed patients. These results did not support the receiver-end deficit hypothesis (NAc deficit -> VTA upregulation) [63]. The accumulating evidence highlights the impairment in VTA-NAc pathway underlying anhedonia in depressed patients, which merits further examination with circuit-based or network-based metrics.

4.3.2 The neural deficits in model-based RL for depression

DS↓. This study found that the RPE_{MB} signal of DS (putamen, caudate) was decreased in the depressed patients, consistent with previous studies [31]. DS tracks action-related task components and has bidirectional functional connectivity to the ventral striatum. The caudate is responsible for representing model-based behavior, and the putamen is responsible for representing model-free behavior [64, 65]. Specifically, the putamen represents and keeps track of the goal selection signal until the next goal is selected, while the caudate mainly encodes the response and outcome information and dynamically maintains the representation of the goal selection signal [66]. Therefore, the results suggest that while encoding RPE_{MF} signals, the representation and maintenance of goals and outcomes may be reduced in the patients with depression, leading to a decline in model-based RL ability.

LOFC↓. Notably, both the RPE_{MB} and RPE_{MF} signals of lOFC were decreased in

the depressed patients. Furthermore, the mediating role of lOFC's RPE_{MB} signal between model-based RL and depressive symptoms was observed. lOFC is involved in both model-based and model-free RPE encoding [67]. OFC encodes a cognitive map of task space which represents the current (especially unobservable) state of the task [51]. LOFC activity is necessary for value updating in a decision-making task which requires model-based control [64]. Moreover, lOFC is involved in salience processing which is sensitive to both positive and negative clues [68]. Therefore, the value updating during model-based as well as model-free RL may be dampened for depressed patients.

MOFC↓. The patients with depression showed decreased RPE_{MB} signal in MOFC, but no abnormalities of RPE_{MF} in MOFC were found, which is inconsistent with previous study [60]. However, this study showed that RPE_{MF} is negatively correlated with depressive symptoms. MOFC is anatomically extensively connected with visceral and motor areas. MOFC is functionally responsible for sensing internal states, encoding the relative value of current actions [69], and associating outcome value with action selection [70]. The results of deficits in RPE_{MB} signal MOFC suggest that depressed patients fail to estimate and update the action value to make model-based decisions.

5 Limitations and directions for future research

The total reward of practice was not taken into consideration to assess the effectiveness of practice. Therefore, there is a lack of objective indexes to validate if the subjects really understood the task. We could not parse overall performance from MF vs. MB strategies. Future studies could include more measurement indicators to assess the degree of task understanding (such as a questionnaire on task structure knowledge, and the comparison between the total scores earned in the practice task) and to parse general performance from RL strategies.

The neuromodulation effects (e.g., dopamine and serotonin) were not assessed and taken into consideration. Previous evidence suggests that elevated dopamine (DA) levels can promote model-based selection behavior [29]. In contrast, reduced DA synthesis and conduction can lead individuals to be more prone to habitual behavior [71]. Serotonin (5-HT) levels regulate the balance between model-based and model-free behavior [72]. Therefore, the individual differences in neurotransmitters may cause an imbalance in the effect of the two types of RL. Future studies need to consider the influences of neurotransmitters underlying deficits in RL processes for depression.

Due to the sample size of the current study, the whole-brain analyses based on family wise error (FWE) correction did not yield significant activation of between-group comparison related to reward prediction error, either model-based or model-free. However, the significant ACC activation of

model-free RPE for the HC group confirmed its role of error detection which drives behavioral switch [73].

Future intervention studies are needed to investigate the cause-effect relationship between RL deficits and depressive/anhedonic symptoms. For example, behavior activation (BA) therapy is recommended by the WHO for supplementary intervention of moderately severe depression, which theoretically aims to engage patients in goal-directed behavior consistent with their values in life [74, 75]. It is still unclear whether BA intervention alters the RL (especially model-based learning) function underlying prefrontal-striatal circuits, which may be promising targets for the personalized treatment of depression. Finally, due to the modest sample size, future studies should replicate the findings to validate the conclusions.

6 Conclusions

(1) Depressed patients showed reward learning (i.e., model-based RL) deficits compared to healthy controls.

(2) Model-based and model-free RL deficits mediated the relationship between stress and anhedonic (or depressive) symptoms for both groups, with underlying specific neural signatures (i.e., RPE_{MF} signals in VTA and caudate).

Consent to Participate declaration

All participants provided written informed consent.

Data Availability declaration

The datasets in the current study are available from the corresponding author on reasonable request.

Funding

This research was financially supported by the Nursery Fund for Young Talents in the Army Medical University (410301053421).

Competing interests

The authors declare that they have no conflict of interest.

Authors' contributions

W.X. and F.Z. design the study, W.X. drafts the manuscript, W.X., H.J. and F.Z. review and revise the manuscript, Z.X. and W.X. recruit the participants, Z.D. performs imaging data acquisition. W.X., Z.Z., H.J. and G.Y. analyze and visualize the data.

Acknowledgements

We would like to express our gratitude to all the volunteers who participated in the study. We also would like to express our sincere gratitude to Dr. Nathanial Daw, Quentin Huys, Bradley Doll, Miriam Sebold, Ross Otto and

Catherine Hartley for their professional guidance in the behavioral and fMRI task design and data modeling; and Dr. Xiaochu Zhang and Zhiyi Chen for their helpful suggestions regarding the research design and preparation of the manuscript.

References:

1. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, Yu Y, Kou C, Xu X, Lu J *et al*: **Prevalence of mental disorders in China: a cross-sectional epidemiological study**. *The Lancet Psychiatry* 2019, **6**(3):211-224.

2. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC: **Prevalence of Depression in the Community from 30 Countries between 1994 and 2014**. *SCI REP-UK* 2018, **8**(1).

3. Rubin DH: **Joy returns last: anhedonia and treatment resistance in depressed adolescents**. *J AM ACAD CHILD PSY* 2012, **51**(4):353-355.

4. Treadway MT, Zald DH: **Reconsidering anhedonia in depression: Lessons from translational neuroscience**. *Neuroscience & Biobehavioral Reviews* 2011, **35**(3):537-555.

5. Thomsen RK: **Measuring anhedonia: impaired ability to pursue, experience, and learn about reward**. *FRONT PSYCHOL* 2015, **6**.

6. Thomsen KR, Whybrow PC, Kringelbach ML: **Reconceptualizing anhedonia: novel perspectives on balancing the pleasure networks in the human brain**. *FRONT BEHAV NEUROSCI* 2015, **9**:49.

7. Bogdan R, Nikolova YS, Pizzagalli DA: **Neurogenetics of depression: a focus on reward processing and stress sensitivity**. *NEUROBIOL DIS* 2013, **52**:12-23.

8. Pizzagalli DA: **Depression, stress, and anhedonia: toward a synthesis and integrated model**. *ANNU REV CLIN PSYCHO* 2014, **10**:393-423.

9. Doya K, Samejima K, Katagiri K, Kawato M: **Multiple Model-Based Reinforcement Learning**. *NEURAL COMPUT* 2002, **14**(6):1347-1369.

10. Bogdan R, Pizzagalli DA: **Acute Stress Reduces Reward Responsiveness: Implications for Depression**. *BIOL PSYCHIAT* 2006, **60**(10):1147-1154.

11. Bogdan R, Santesso DL, Fagerness J, Perlis RH, Pizzagalli DA: **Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning**. *J NEUROSCI* 2011, **31**(37):13246-13254.

12. Vrieze E, Pizzagalli DA, Demyttenaere K, Hompes T, Sienaert P, de Boer P, Schmidt M, Claes S: **Reduced Reward Learning Predicts Outcome in Major Depressive Disorder**. *BIOL PSYCHIAT* 2013, **73**(7):639-645.

13. Pizzagalli DA, Bogdan R, Ratner KG, Jahn AL: **Increased perceived stress is associated with blunted hedonic capacity: Potential implications for depression research**. *BEHAV RES THER* 2007, **45**(11):2742-2753.

14. Daw ND, Dayan P: **The algorithmic anatomy of model-based evaluation**. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2014, **369**(1655):20130478.

15. Doll BB, Simon DA, Daw ND: **The ubiquity of model-based reinforcement learning**. *CURR OPIN NEUROBIOL* 2012, **22**(6):1075-1081.

16. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ: **Model-Based Influences on Humans' Choices and Striatal Prediction Errors**. *NEURON* 2011, **69**(6):1204-1215.

17. Maddox WT, Chandrasekaran B, Smayda K, Yi H, Koslov S, Beevers CG: **Elevated depressive symptoms enhance reflexive but not reflective auditory category learning**. *CORTEX* 2014, **58**:186-198.

18. Blanco NJ, Otto AR, Maddox WT, Beevers CG, Love BC: **The influence of depression symptoms on exploratory decision-making**. *COGNITION* 2013, **129**(3):563-568.

19. Schwabe L, Wolf OT: **Stress-induced modulation of instrumental behavior: from goal-directed to habitual control of action**. *BEHAV BRAIN RES* 2011, **219**(2):321-328.

20. Radenbach C, Reiter AMF, Engert V, Sjoerds Z, Villringer A, Heinze H, Deserno L, Schlagenhauf F: **The interaction of acute and chronic stress impairs model-based behavioral control**. *PSYCHONEUROENDOCRINO* 2015, **53**:268-280.

21. Braun S, Hauber W: **Acute stressor effects on goal-directed action in rats**. *LEARN MEMORY* 2013, **20**(12):700-709.

22. Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND: **Working-memory capacity protects model-based learning from stress**. *Proceedings of the National Academy of Sciences* 2013, **110**(52):20941-20946.

23. Heller AS, Ezie CEC, Otto AR, Timpano KR: **Model-based learning and individual differences in depression: The moderating role of stress**. *BEHAV RES THER* 2018, **111**:19-26.

24. Patterson E, Hunter C, Gray Z, Trudell E, Shields G: **Testing the theory of stress as a cumulative prediction error**. *PSYCHONEUROENDOCRINO* 2023, **153**:106214.

25. Knolle F, Sen P, Culbreth A, Koch K, Schmitz-Koep B, Gürsel DA, Wunderlich K, Avram M, Berberich G, Sorg C *et al*: **Investigating disorder-specific and transdiagnostic alterations in model-based and model-free decision-making**. *Journal of Psychiatry and Neuroscience* 2024, **49**(6):E388-E401.

26. den Ouden HEM, Kok P, de Lange FP: **How Prediction Errors Shape Perception, Attention, and Motivation**. *FRONT PSYCHOL* 2012,

3:548.

27. Schultz W: **Dopamine reward prediction-error signalling: a two-component response**. *NAT REV NEUROSCI* 2016, **17**(3):183-195.

28. Smittenaar P, FitzGerald THB, Romei V, Wright ND, Dolan RJ: **Disruption of Dorsolateral Prefrontal Cortex Decreases Model-Based in Favor of Model-free Control in Humans**. *NEURON* 2013, **80**(4):914-919.

29. Wunderlich K, Smittenaar P, Dolan RJ: **Dopamine Enhances Model-Based over Model-Free Choice Behavior**. *NEURON* 2012, **75**(3):418-424.

30. Dombrovski AY, Szanto K, Clark L, Reynolds CF, Siegle GJ: **Reward signals, attempted suicide, and impulsivity in late-life depression**. *JAMA PSYCHIAT* 2013, **70**(10):1.

31. Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M, Reid I, Hall J, Steele JD: **Expected value and prediction error abnormalities in depression and schizophrenia**. *BRAIN* 2011, **134**(6):1751-1764.

32. Kumar P, Waiter G, Ahearn T, Milders M, Reid I, Steele JD: **Abnormal temporal difference reward-learning signals in major depression**. *BRAIN* 2008, **131**(Pt 8):2084-2093.

33. Villano WJ, Heller AS: **Depression is associated with blunted affective responses to naturalistic reward prediction errors**. *PSYCHOL MED* 2024, **54**(9):1956-1964.

34. Heo S, Sung Y, Lee SW: **Effects of subclinical depression on prefrontal–striatal model-based and model-free learning**. *PLOS COMPUT BIOL* 2021, **17**(5):e1009003.

35. Cremer A, Kalbe F, Gläscher J, Schwabe L: **Stress reduces both model-based and model-free neural computations during flexible**

learning. *NEUROIMAGE* 2021, **229**:117747.

36. Liu Q, Ely BA, Schwartz JJ, Alonso CM, Stern ER, Gabbay V: **Reward function as an outcome predictor in youth with mood and anxiety symptoms**. *J AFFECT DISORDERS* 2021, **278**:433-442.

37. Ely BA, Liu Q, DeWitt SJ, Mehra LM, Alonso CM, Gabbay V: **Data-driven parcellation and graph theory analyses to study adolescent mood and anxiety symptoms**. *TRANSL PSYCHIAT* 2021, **11**(1):266.

38. Beck AT Brown GK SRA: **Manual of Beck Depression Inventory-II**. *American University Washington Dc* 1996, **21**(88).

39. Zhen W, Chengmei Y, Jia H, Zezhi L, Jue C, Haiyin Z, Yiru F, Zeping X: **The reliability and validity of the beck depression scale (BDI-II) in Chinese depressed patients**. *Chinese Mental Health Journal* 2011, **25**(6):476-480.

40. Clark LA, Watson D: **Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications**. *J ABNORM PSYCHOL* 1991, **100**(3):316-336.

41. Yang XL, Yao SQ: **Reliability and Validity of the Chinese Version of the Mood and Anxiety Symptoms Questionnaire for University Students**. *Chinese Journal of Clinical Psychology* 2009.

42. Cohen S, Kamarck T, Mermelstein R: **A global measure of perceived stress**. *J HEALTH SOC BEHAV* 1983, **24**(4):385-396.

43. Wang Z, Wang Y, Wu Z, Chen D, Chen J, Xiao Z: **Reliability and validity of the Chinese version of Perceived Stress Scale**. *Journal of Shanghai Jiao Tong University Medical Science* 2015, **35**(10):1448-1451.

44. Decker JH, Otto AR, Daw ND, Hartley CA: **From Creatures of Habit to Goal-Directed Learners**. *PSYCHOL SCI* 2016, **27**(6):848-858.

45. Culbreth AJ, Westbrook A, Daw ND, Botvinick M, Barch DM: **Reduced model-based decision-making in schizophrenia.** *J ABNORM PSYCHOL* 2016, **125**(6):777-787.

46. Gillan CM, Vaghi MM, Hezemans FH, van Ghesel Grothe S, Dafflon J, Brühl AB, Savulich G, Robbins TW: **Experimentally induced and real-world anxiety have no demonstrable effect on goal-directed behaviour**. *PSYCHOL MED* 2020:1-12.

47. Huys QJM: **Chapter 10 - Bayesian Approaches to Learning and Decision-Making**. In *Computational Psychiatry*. Edited by Anticevic A, Murray JD: Academic Press; 2018:247-271.

48. Hayes AF: **PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling**.; 2012.

49. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ: **Bayesian model selection for group studies**. *NEUROIMAGE* 2009, **46**(4):1004-1017.

50. Lieder F, Goodman N, Huys Q: **Learned helplessness and generalization**. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Berlin, Germany: Cognitive Science Society; 2013:900-905.

51. Schuck NW, Cai MB, Wilson RC, Niv Y: **Human Orbitofrontal Cortex Represents a Cognitive Map of State Space**. *NEURON* 2016, **91**(6):1402-1412.

52. Xu P, Wang K, Lu C, Dong L, Chen Y, Wang Q, Shi Z, Yang Y, Chen S, Liu X: **Effects of the chronic restraint stress induced depression on reward-related learning in rats**. *BEHAV BRAIN RES* 2017, **321**:185-192.

53. McEwen BS, Morrison JH: **The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course**. *NEURON* 2013, **79**(1):16-29.

54. Schwabe L, Hoffken O, Tegenthoff M, Wolf OT: **Stress-induced enhancement of response inhibition depends on mineralocorticoid receptor activation**. *PSYCHONEUROENDOCRINO* 2013, **38**(10):2319-2326.

55. MacAulay RK, McGovern JE, Cohen AS: **Understanding Anhedonia: The Role of Perceived Control**. In *Anhedonia: A Comprehensive Handbook Volume I: Conceptual Issues And Neurobiological Advances*. Edited by Ritsner MS. Dordrecht: Springer Netherlands; 2014:23-49.

56. Maier SF, Seligman MEP: **Learned helplessness at fifty: Insights from neuroscience.** *PSYCHOL REV* 2016, **123**(4):349-367.

57. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control**. *NAT NEUROSCI* 2005, **8**(12):1704-1711.

58. Gläscher J, Daw N, Dayan P, O'Doherty JP: **States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning**. *NEURON* 2010, **66**(4):585-595.

59. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ: **Dissociable roles of ventral and dorsal striatum in instrumental conditioning**. *SCIENCE* 2004, **304**(5669):452-454.

60. Rothkirch M, Tonn J, Köhler S, Sterzer P: **Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder**. *BRAIN* 2017, **140**(4):1147-1157.

61. Greenberg T, Chase HW, Almeida JR, Stiffler R, Zevallos CR, Aslam HA, Deckersbach T, Weyandt S, Cooper C, Toups M *et al*: **Moderation of the Relationship Between Reward Expectancy and Prediction Error-Related Ventral Striatal Reactivity by Anhedonia in Unmedicated Major Depressive Disorder: Findings From the EMBARC Study**. *AM J PSYCHIAT* 2015, **172**(9):881-891.

62. Boureau YL, Dayan P: **Opponency revisited: competition and**

cooperation between dopamine and serotonin. *NEUROPSYCHOPHARMACOL* 2011, **36**(1):74-97.

63. Chen C, Takahashi T, Nakagawa S, Inoue T, Kusumi I: **Reinforcement learning in depression: A review of computational research**. *Neuroscience & Biobehavioral Reviews* 2015, **55**(0):247-267.

64. Gremel CM, Costa RM: **Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions**. *NAT COMMUN* 2013, **4**:2264.

65. Kim H, Lee D, Jung MW: **Signals for Previous Goal Choice Persist in the Dorsomedial, but Not Dorsolateral Striatum of Rats**. *The Journal of Neuroscience* 2013, **33**(1):52-63.

66. McNamee D, Liljeholm M, Zika O, O'Doherty JP: **Characterizing the Associative Content of Brain Structures Involved in Habitual and Goal-Directed Actions in Humans: A Multivariate fMRI Study**. *The Journal of Neuroscience* 2015, **35**(9):3764-3771.

67. Nasser HM, Calu DJ, Schoenbaum G, Sharpe MJ: **The Dopamine Prediction Error: Contributions to Associative Models of Reward Learning**. *FRONT PSYCHOL* 2017, **8**.

68. Nusslock R, Alloy LB: **Reward processing and mood-related symptoms: An RDoC and translational neuroscience perspective**. *J AFFECT DISORDERS* 2017, **216**:3-16.

69. McDannald MA, Jones JL, Takahashi YK, Schoenbaum G: **Learning theory: a driving force in understanding orbitofrontal function**. *NEUROBIOL LEARN MEM* 2014, **108**:22-27.

70. Strait CE, Blanchard TC, Hayden BY: **Reward value comparison via mutual inhibition in ventromedial prefrontal cortex**. *NEURON* 2014, **82**(6):1357-1366.

71. de Wit S, Standing HR, DeVito EE, Robinson OJ, Ridderinkhof KR, Robbins TW, Sahakian BJ: **Reliance on habits at the expense of goal-directed control following dopamine precursor depletion**. *PSYCHOPHARMACOLOGY* 2012, **219**(2):621-631.

72. Sanchez CL, Biskup CS, Herpertz S, Gaber TJ, Kuhn CM, Hood SH, Zepf FD: **The Role of Serotonin (5-HT) in Behavioral Control: Findings from Animal Research and Clinical Implications**. *INT J NEUROPSYCHOPH* 2015, **18**(10):pyv50.

73. Cole N, Harvey M, Myers-Joseph D, Gilra A, Khan AG: **Prediction-error signals in anterior cingulate cortex drive task-switching**. *NAT COMMUN* 2024, **15**(1):7088.

74. Martell CR, Dimidjian S, Herman-Dunn R: **Behavioral activation for depression: a clinician's guide**. *Dunn* 2010(8):837-838.

75. Lejuez CW, Hopko DR, Acierno R, Daughters SB, Pagoto SL: **Ten Year Revision of the Brief Behavioral Activation Treatment for Depression: Revised Treatment Manual**. *BEHAV MODIF* 2011, **35**(2):111-161.