*Psychological Medicine*

**cambridge.org/psm**

# Factual and counterfactual learning in major adolescent depressive disorder, evidence from an instrumental learning study

Qiang Shen[1,2,3,*], Shiguang Fu[1,2,3,*], Xiaoying Jiang[4], Xiaoyu Huang[4], Doudou Lin[5], Qingyan Xiao[1,2,3], Sitti Khadijah[5], Yaping Yan[6], Xiaoxing Xiong[7], Jia Jin[1,2,3], Richard P. Ebstein[8], Ting Xu[9], Yiquan Wang[4] and Jun Feng[10]

[1]Shanghai Key Laboratory of Brain-Machine Intelligence for Information Behavior (Ministry of Education), 201620, Shanghai, China; [2]School of Business and Management, Shanghai International Studies University, 201620, Shanghai, China; [3]Joint Lab of Finance and Business Intelligence, Guangdong Institute of Intelligence Science and Technology, 519031, Zhuhai, China; [4]Hangzhou Mental Health Center of Children and Adolescents, Hangzhou Seventh People's Hospital, 310006, Hangzhou, China; [5]School of Management, Zhejiang University of Technology, 310023, Hangzhou, China; [6]Department of Neurology, The Second Affiliated Hospital of Zhejiang University, 310009, Hangzhou, China; [7]Department of Neurosurgery, Renmin Hospital of Wuhan University, 430060, Wuhan, China; [8]China Center for Behavioral Economics and Finance, Southwestern University of Finance & Economics, 611130, Chengdu, China; [9]School of Business, University of Ningbo, 315210, Ningbo, China and [10]School of Economics, Hefei University of Technology, 230601, Hefei, China

## Abstract

**Background.** The incidence of adolescent depressive disorder is globally skyrocketing in recent decades, albeit the causes and the decision deficits depression incurs has yet to be well-examined. With an instrumental learning task, the aim of the current study is to investigate the extent to which learning behavior deviates from that observed in healthy adolescent controls and track the underlying mechanistic channel for such a deviation.

**Methods.** We recruited a group of adolescents with major depression and age-matched healthy control subjects to carry out the learning task with either gain or loss outcome and applied a reinforcement learning model that dissociates valence (positive *v.* negative) of reward prediction error and selection (chosen *v.* unchosen).

**Results.** The results demonstrated that adolescent depressive patients performed significantly less well than the control group. Learning rates suggested that the optimistic bias that overall characterizes healthy adolescent subjects was absent for the depressive adolescent patients. Moreover, depressed adolescents exhibited an increased pessimistic bias for the counterfactual outcome. Lastly, individual difference analysis suggested that these observed biases, which significantly deviated from that observed in normal controls, were linked with the severity of depressive symoptoms as measured by HAMD scores.

**Conclusions.** By leveraging an incentivized instrumental learning task with computational modeling within a reinforcement learning framework, the current study reveals a mechanistic decision-making deficit in adolescent depressive disorder. These findings, which have implications for the identification of behavioral markers in depression, could support the clinical evaluation, including both diagnosis and prognosis of this disorder.

## Original Article

*These authors contributed equally to this work.

**Cite this article:** Shen Q *et al* (2023). Factual and counterfactual learning in major adolescent depressive disorder, evidence from an instrumental learning study. *Psychological Medicine* 1–11. https://doi.org/10.1017/S0033291723001307

Received: 31 January 2022
Revised: 10 March 2023
Accepted: 17 April 2023

**Keywords:**
Choice bias; depression; reinforcement learning; reward prediction error

**Corresponding author:**
Yiquan Wang;
Email: wangyiquan1978@126.com;
Jun Feng;
Email: cb8226@hotmail.com

## Factual and counterfactual learning in adolescent depressive disorder, evidence from an instrumental learning study

In the modern era, depression has increasingly emerged as one of the most challenging and troubling of mental health disorders, apparently exacerbated as a consequence of the COVID-19 pandemic (Miller & Campo, 2021; Santomauro et al., 2021). During development, the incidence of depressive disorder rises strikingly at puberty and following adult trends, is more pronounced for females (Paus, Keshavan, & Giedd, 2008; Stevanovic, Jancic, & Lakic, 2011). Notably, both in Asia and the West, the prevalence of adolescent depression is skyrocketing in recent years and has emerged as a major public health concern of the first order. Notably, adolescent depression is gaining considerable attention both from governments and society as a whole (Clayborne, Varin, & Colman, 2019; Lu, 2019; Twenge, Cooper, Joiner, Duffy, & Binau, 2019).

Prior studies of depression, including in adolescence, suggest that individuals with depressive symptoms experience the world around them in a more negative manner which presents either as hypersensitivity toward punishment or hyposensitivity to rewards (Kube, Schwarting, Rozenkrantz, Glombiewski, & Rief, 2020; Nielson et al., 2021). Toward a better understanding of the latent mechanisms underpinning adolescent depression, a considerable group of studies

have applied instrumental learning tasks to reveal behavioral regularities of decision making (Bavard, Rustichini, & Palminteri, 2021; Frank, Seeberger, & O'reilly, 2004; Gillan, Otto, Phelps, & Daw, 2015). The key feature of the instrumental reinforcement learning task is that it connects situations with actions and individuals to achieve goal-oriented behaviors through trial-and-error exploration (Sutton & Barto, 2018). Notably, action selection will not only determine imminent rewards, but could also shape belief and value formation in subsequent trials. These considerations make it feasible to characterize both the qualitative and quantitative dynamics of learning behavior in both adult and adolescent depression.

Computational modeling of reward (Berwian et al., 2020; Montague, Dolan, Friston, & Dayan, 2012), and the refinement of our understanding of reward as 'reward prediction error' (RPE), especially as encoded by midbrain dopamine neurons, offers a window into a better understanding of the blunted reward response in the adolescent depressive patient (Ng, Alloy, & Smith, 2019; Stringaris et al., 2015). In the framework of RPE, it is crucial to differentiate the positive RPE from the negative RPE and track the potential asymmetric response over the valence (either positive or negative). Implementing this approach makes it feasible to examine the potential asymmetric response, viz. whether the valence reflects 'good news' or 'bad news' for belief updating. As indicated in widely-used psychological paradigms implemented in recent series of studies, instead of arriving at an accurate belief, individual tends to interpret the choice-contingent outcome to form a belief to achieve a desirable manner for their own sake (e.g. optimistic bias, Bromberg-Martin & Sharot, 2020; Ma et al., 2016; Sharot, 2011; Sharot, Riccardi, Raio, & Phelps, 2007).

Applying the instrumental learning task, for typical individuals, Lefebvre, Lebreton, Meyniel, Bourgeois-Gironde, and Palminteri (2017) suggests that there is an inclination to assign a higher weight (learning rate) toward the reward, or positive prediction error, in comparison to negative prediction error. This tendency has been termed optimistic bias and this optimistic update is congruent at the neural level, as revealed by fMRI analysis, with an increasing RPE signal in the reward-related region brain regions (striatum). This optimistic bias is coincident with the findings from the studies of belief updating with a priori desirability or undesirability (Sharot & Garrett, 2016; Sharot, Velasquez, & Dolan, 2010).

In depressive disorders, however, Korn, Sharot, Walter, Heekeren, and Dolan (2014) found that such an optimism bias with prior belief tends to be absent for depressive patients and, moreover, is correlated with a higher Beck Depression Inventory score. From the perspective of RPE, Kumar et al. (2018) suggests the depressive subjects selectively reduce positive RPE rather than enhance punishment or negative RPE. At the neural level, depressive subjects exhibit corresponding reduced striatal-midbrain connectivity. Altogether, the findings from RPE (positive *v.* negative) suggest that there might be a selective and asymmetric impairment of reward related processing in the depressive subjects.

Interestingly, recent studies also have tracked how choice shapes belief formation and behavioral adjustment in learning tasks. In the field of economics, Hartzmark, Hirshman, and Imas (2021) reported that the exogenously manipulated ownership could elicit the optimistic belief from positive information. In psychology, Palminteri, Lefebvre, Kilford, and Blakemore (2017) and Chambon et al. (2020) found that the self-determined choice itself could modulate the confirmation of the chosen choice (factual learning) and disconfirm the unchosen option (counterfactual learning, see also Lefebvre, Summerfield, & Bogacz, 2021; Tarantola, Folke, Boldt, Perez, & De Martino, 2021 for recent progress). Importantly, empirical studies suggest that adolescent individuals might fail to benefit from counterfactual information in comparison to adult subjects (Palminteri, Kilford, Coricelli, & Blakemore, 2016). Despite the importance of this phenomenon, few studies have yet examined the contours of behavioral regularities for subjects with depressive disorder from such a learning perspective nor attempted to unravel how such patients use information from counterfactual observation.

Last but not least, recent studies have attempted to tease apart the underlying mechanism of distorted belief updating in depression through the lens of intra-trial dynamics of choice implementation within the context of reinforcement learning. For instance, Pedersen, Frank, and Biele (2017) and Fontanesi, Gluth, Spektor, and Rieskamp (2019) have introduced the measurement of response time to value-based decision making in a reinforcement learning framework. With respect to adolescent depressive disorder, previous studies already suggested that there is a prolonged response time of the adolescent depressive subjects compared with those of the healthy controls (Chase et al., 2010). Nevertheless, few studies have integrated this potentially crucial component into computational modeling toward a more complete understanding of reinforcement learning for psychiatric diseases (see Wiehler, Chakroun, & Peters, 2021 for a most recent attempt for gambling disorders).

To our knowledge, no existing studies have examined how choice, prediction error and intra-trial dynamics integrally shape the learning processing in adolescent depressive disorders. Toward capturing the learning characteristics of adolescent depression, we employ a two-bandit instrumental learning task (following Palminteri et al., 2017), which involves both partial and complete feedback in adolescent patients and age matched normal adolescent subjects. By leveraging the advantages of using computational modeling, and including response time which likely reflects salient aspects of the decision process, this design allows us to evaluate the extent to which adolescent depressive subjects make use of choice-related outcome (factual *v.* counterfactual), RPE (positive *v.* negative), and their interaction, which we hypothesize shape dysfunctional decision making in such patients.

## Methods

### Subjects

We recruited 84 adolescent subjects (age from 12 to 18), including 42 adolescent subjects with major depression (MDD) (age: mean = 15.19, S.D. = 1.73; 4 males) and 42 healthy subjects (age: mean = 14.88, S.D. = 2.00; 13 males). Subjects diagnosed with major depressive disorders were hospitalized patients at the Seventh People's Hospital of Hangzhou, Zhejiang Province in a specialized ward for adolescent psychiatric patients. Subjects in the control group were students from local junior and senior high schools in Hangzhou. All adolescent depressive patients were diagnosed by board-qualified psychiatrists and patients were characterized with Hamilton Depression Rating Scale (HAMD, 24-item version, Zhang & He, 2015) scores equal or greater than 20 (mean = 31.69, S.D. = 6.55, see online Supplementary Table S1 for the detailed clinical status including medication, comorbidity etc., of the

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

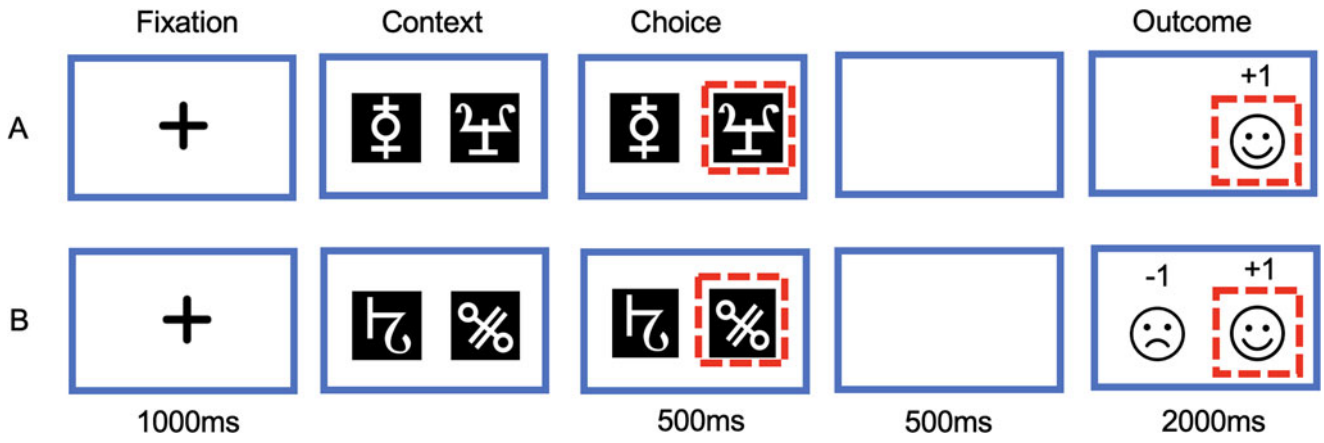


**Figure 1.** A trial starts with a 'cross' sign and two abstract symbols present on the screen which would keep fixed across one whole block (24 trials), the chosen option is highlighted with a colored rectangle once the subjects selected it with the key press and then the chosen outcome is revealed only for the partial condition (Panel A) and the outcome of the unchosen option is also revealed for complete option (Panel B). The winning outcome is 1 point together with a smiley face while the loss outcome is −1 point with a sad face. The earned money is accumulated for all earned points across 288 trials for the final performance score.

MDD patients). The written informed consent was obtained from all subjects prior to their participation. The study was officially approved by the Ethics committee of the Seventh People's Hospital in Hangzhou, Zhejiang province (No. 2020029).

To rule out potential confounds related to cognitive abilities, all subjects were asked to complete part of the Raven's Standard Progressive Matrices (set E, Raven & Court, 1998); the average score for the MDD group was 6.38 (S.D. = 2.06) and compared to 6.69 (S.D. = 1.75) for healthy controls ($t(79) = 0.723$, $p = 0.472$).

### Behavioral task

Following the original design of Palminteri et al. (2017), the subjects were instructed to carry out a probabilistic instrumental task and aimed to earn monetary payoff through trial and error. In a series of consecutive 24 trials within each block, the subjects were asked to make a choice selection out of two fixed abstract symbols (Fig. 1 for a demo). Adopting a within subject strategy, we involved two factors: *information* (partial and complete) and *probability* (3 levels: symmetric, asymmetric and reversal) for the learning task. For the partial condition, only the choice-contingent outcome was revealed whereas for the complete condition, both the chosen and unchosen option-related outcome was presented at the end of the trial. The outcome was either winning one point (+1) for gain or losing one point (−1) for loss. For the 'Symmetric' condition, both options had a half-half chance of getting a reward or incurring a loss; with respect to the 'Asymmetric' condition, one symbol was linked with a ¾ chance of gain, whereas the other option was coupled with a ¼ chance of winning; In the final 'Reversal' condition, where one option was associated with an 83% chance of getting a reward and the other option was linked with a 17% chance of getting a reward for the 1st half of the total 24 trials, whereas in the 2nd half of the block, the reward contingencies for two options were reversed. Although we fixed that the winning probabilities in the two options added up to 1, in the experiment, gain or loss is independently determined for each option. Within each block, for the fixed pair of abstract symbols, the subjects need to discover the regularities gradually without any prior informed knowledge. There were two blocks for all the 6 levels (2×3) as elaborated above, resulting in 288 trials for each participant. A schematic illustration of a complete trial was exhibited in Fig. 1. The task was programmed with open-source package PsychoPy v3 (version 3.2.4, https://www.psychopy.org/) on the python platform. The abstract symbols were reproduced or newly created by a professional designer with high resolution, here we thank Dr. Stefano Palminteri for providing us with their initial symbol library from their previous publication for reference (Palminteri et al., 2017). To minimize a potential order effect, the task was counterbalanced at the condition level across subjects; Within each subject, the order of the two designated options with predetermined probability (symmetric, asymmetric and reversal) was randomly determined at the very start of the block and then the position keep unchanged across the whole block (24 trials). As illustrated in Fig. 1, once the subjects used the numeric keypad to select their preferred option (key '4' for the left option and key '6' for the right option), the chosen option was highlighted immediately and subsequently the outcome of the chosen option was revealed for the partial condition (panel A). For the complete condition, both the outcome of the chosen and the unchosen alternative were shown to the subjects concurrently (panel B). At the end of the experimental task, the subjects got their earned payoff according to their realistic performance as well as their show-up fee and the performance score was converted into compensation with a fixed exchange rate (3:1) in the form of payoff matched gifts.

### Behavioral data analysis

#### Payoff and choice accuracy

For the initial step, we calculated the aggregate payoff as well as the frequency of correct choice selection of both the depressive disorder patients and their age-matched normal adolescent subjects. As there was no theoretically correct option for the symmetric condition, we only counted the values for the asymmetric and reversal condition. Notably, the correct option was reversed for the 2nd half of the block in the reversal condition, therefore, we analyzed the correct choice rate separately for these two learning phases (Palminteri et al., 2017). To illustrate the dynamic learning process of the implemented task, we graphically show the cumulative frequency of choice accuracy across 24 trials (Fig. 2).

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

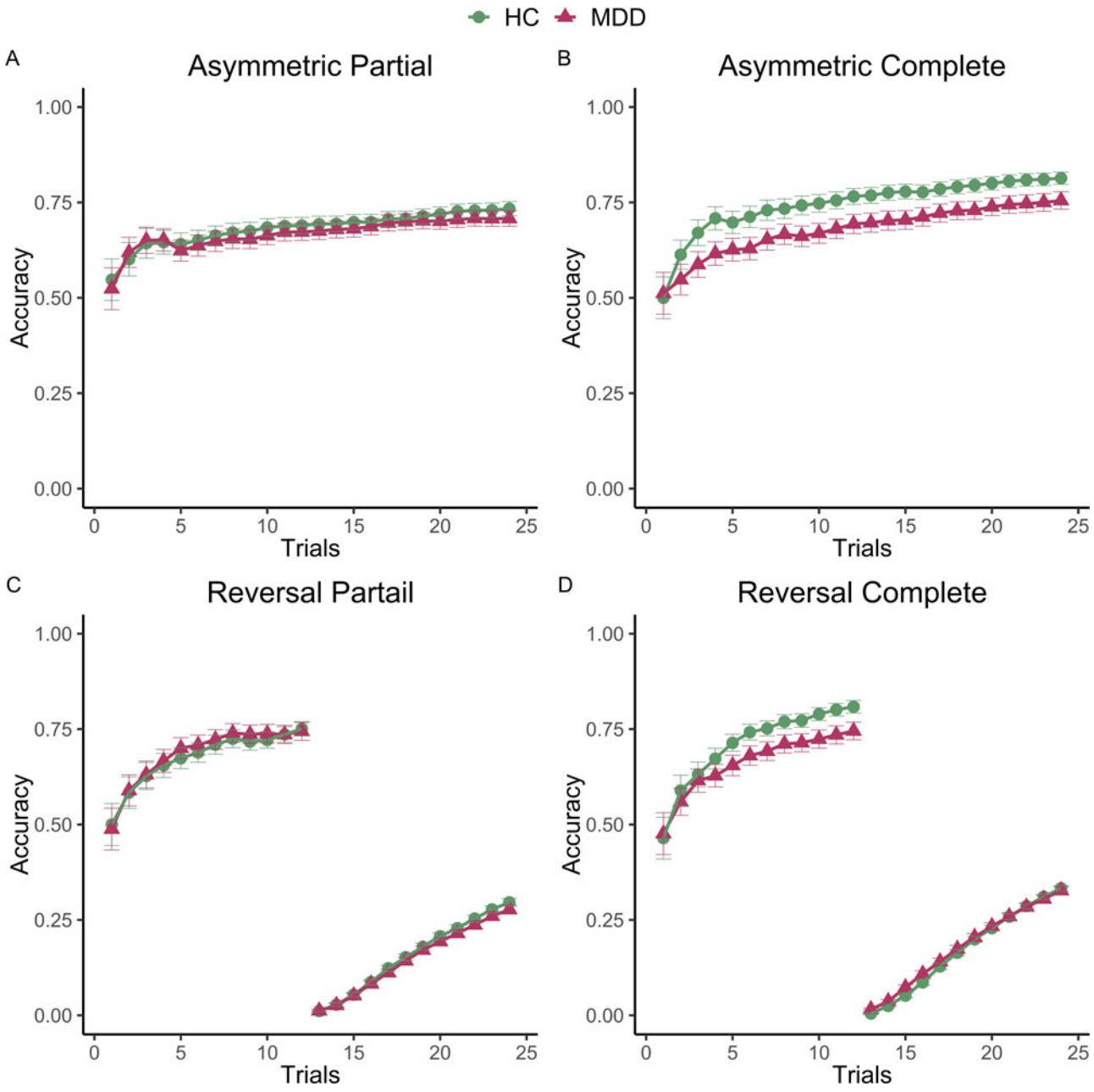


**Figure 2.** It illustrates the trial-wise cumulative choice accuracy for asymmetric (top panel) and reversal condition (bottom panel) across 24 trials in each block. The left panel represents the partial condition and the right panel is for the complete condition. Note that there is no definition of correct choice for the symmetric condition, therefore it is omitted in the figure.

For the statistical inference for the payoff, we carry out a multilevel linear regression, DV $payoff_{it}$ is subject $i$'s at trial $t$ (assigns as 1 if the current feedback is reward and −1 otherwise). The key independent variables relating to the experiment are *Group* (equal to 1 for healthy control and 0 for MDD), *Information* (equal to 1 for partial and 0 for complete feedback) and *probability* (dummy variables: Symmetric, Asymmetric and Reversal, for the dummy example using Reversal as a reference 0). Given the within-subject serial correlation across trials, the standard error was clustered at the individual level. The analysis applies for the choice accuracy is in a similar spirit as implemented for the payoff. The differences lie in two aspects, 1: the dependent variable is $Choice_{it}$ (assigns as 1 if the choice is accurate and 0 otherwise, hence the logistic regression is applied); 2: here we only focused on the asymmetric and reversal conditions as there is no explicit definition of the correct answer for symmetric condition.

### *Response time (RT)*

Although reinforcement learning task primarily focuses on the choice data, the response time also indexes relevant information concerning both choice and decision values (Palminteri et al., 2016) including recent advances over the RL drift diffusion model (e.g. Pedersen & Frank, 2020). Therefore, as a computational psychiatric investigation, it is of apparent interest to examine the potential link between the duration of response time and the depressive disorders. Given the skewed distribution of the RT

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

data by nature, we firstly removed the RTs outliers when they were above the 0.75 quartile by more than 1.5 times the interquartile range, or below the 0.25 quantile by more than 1.5 times the interquartile range. We then perform the *Box-Cox* transformation of the RT and run the multi-level linear regression in a similar manner as that carried out for payoff (section Payoff and choice accuracy).

### Reinforcement learning model with standard random utility specification (Model I)

With the classical Rescorla-Wagner (RW) model, a standard Q-learning algorithm is applied to generate trial-by-trial estimates of Q-values and prediction errors (PEs). To test the potential asymmetric effect toward the valence of the prediction error (positive *v.* negative), we separated the positive RPE and negative RPE based on the realized magnitude of the trial-wise PE. Moreover, there were two different kinds of feedback, chosen outcome ($R_c$) *v.* unchosen outcome ($R_u$) for the complete condition, which results in both factual and counterfactual learning for subjects in such a context. Therefore, as illustrated in equations 1 and 3 (Palminteri et al., 2016), we derived the reinforcement learning model with two separate learning rates for the partial condition ($\alpha_+$, $\alpha_-$) and four learning rates for the complete condition ($\alpha_{c+}$, $\alpha_{c-}$, $\alpha_{u+}$, $\alpha_{u-}$). As illustrated in Fig. 2, there was no explicit definition of correct choice for the symmetric condition which might also lead to the uninformative feature of the learning rate derived from the model. Moreover, the RL model analysis we applied here might not be sufficient for the reversal task, as a Rescorla – Wagner model with associability-gated learning rate normally requires longer trials (e.g. 60 trials), which could be used to test the cognitive flexibility of the belief updating (see Mukherjee, Filipowicz, Vo, Satterthwaite, & Kable, 2020; Raio, Hartley, Orederu, Li, & Phelps, 2017 for reference). Therefore, we mainly focused on the asymmetric condition for the subsequent learning rate analysis (see online Supplementary Fig. S3-S4 for the results of the symmetric and reversal condition). For the two options in each block, the value of $Q$ was assigned as 0 for the initial trial. For those trials $t > 1$, the $Q$-value of the chosen option is updated according to the following rule (factual learning module):

$$Q_c(t+1) = Q_c(t) + \begin{array}{ll} \alpha_{c+} * PE_c(t), & if\ PE_c(t) > 0 \\ \alpha_{c-} * PE_c(t), & if\ PE_c(t) < 0 \end{array} \quad (1)$$

In the 1st equation, $\alpha_c$ is divided into $\alpha_{c+}$ and $\alpha_{c-}$ according to $PE_c(t)$. $PE_c(t)$ is the prediction error of the chosen option, which is calculated as:

$$PE_c(t) = R_c(t) - Q_c(t) \quad (2)$$

$R_c(t)$ is the reward outcome of the chosen option at current trial $t$. $R_c(t) - Q_c(t)$ represents chosen RPE at current trial $t$.

For the complete information condition, in addition to the chosen option, the value of the unchosen option is also updated according to the following rule (counterfactual learning module):

$$Q_u(t+1) = Q_u(t) + \begin{array}{ll} \alpha_{u+} * PE_u(t), & if\ PE_u(t) > 0 \\ \alpha_{u-} * PE_u(t), & if\ PE_u(t) < 0 \end{array} \quad (3)$$

In equation 3, according to $PE_u(t)$, $\alpha_u$ is divided into $\alpha_{u+}$ and $\alpha_{u-}$. $PE_u(t)$ is the prediction error of the unchosen option, which is calculated as:

$$PE_u(t) = R_u(t) - Q_u(t) \quad (4)$$

$R_u(t)$ is the unchosen reward outcome on the current trial $t$. $R_u(t) - Q_u(t)$ represents unchosen RPE on the current trial $t$.

The probability of an individual's actual choice at trial $t$ is estimated on the basis of the *softmax* rule:

$$P_c(t) = \frac{e^{(Q_c(t)*\beta)}}{(e^{(Q_c(t)*\beta)} + e^{(Q_u(t)*\beta)})} \quad (5)$$

$$LL = \log(P(Data|Model)) \quad (6)$$

For equation 5, $P_c(t)$ is the probability of choosing the option at current trial $t$. $Q_c(t)$ and $Q_u(t)$ is the updated value at the current trial. $\beta$ is referred to the inverse temperature parameter that adjusts the stochasticity of decision-making. Maximum Likelihood Estimation (MLE) is implemented to estimate the parameters with software R with the self-written scripts and the R package *DEoptim* (Mullen, Ardia, Gil, Windover, & Cline, 2011) is used to achieve a global optimization for parameter estimation. Negative log-likelihoods (LL) are used to compute classical model selection criteria (Equation 6, see online Supplementary Table S7 for details).

### Reinforcement learning model with response time (Model II)

Although with the dynamic update feature of the $Q_t$ of each option as the trial evolves in each block, the Q-learning model illustrated above still falls within in the framework of the random utility model (RUM, McFadden, 1973). As highlighted recently by Webb (2019), with the omission of the endogenous variable RT which is closely correlated with the utility difference between options, might lead to misspecification of the choice probabilities and bias estimates of model parameters. Therefore, we drew on the method proposed by his work (Webb, 2019) and involved the response time into the RL model for evidence accumulation (EA) consideration. Critically, we specified the choice probabilities

$$P_c(t) = \frac{e^{(Q_c(t)*\beta)}}{(e^{(Q_c(t)*\beta)} + e^{(Q_u(t)*\beta)})}$$

with $\beta = e^{s+g*(RT_t - RT_{mean})}$. $s + g*(RT_t - RT_{mean})$ is a linear form where $RT_t$ is the response time at trial $t$ and $RT_{mean}$ is the average response time for certain condition. Therefore, $RT_t - RT_{mean}$ can be regarded as the centered response time, and $s$ and $g$ are two parameters in the model (see online Supplementary Table S8). According to Webb (2019), the response time is negatively correlated to the absolute difference between the two options. Generally speaking, the choice noise is less when the absolute difference between the two options is larger. Therefore, at the aggregate level, parameter $g$ should be negative. This reasoning is consistent with our estimation results (see online Supplementary Table S8).

We also conduct a model selection procedure to compare the standard model to the model with response time (see online Supplementary Model selection). The result of model selection shows that introducing response time to the standard model cannot achieve a significant better explanatory power. However, as Webb

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

(2019) argued, the main improvement of including response time is to reduce the bias of estimates. To further verify this argument within the framework of reinforcement learning which inherently contains dynamic updating, we carry out a simulation in which we simulate the data including both the choices and response times (Miletić et al., 2021). Then, we recover the parameters using the two models discussed in this paper (see online Supplementary Simulation recovery). We find that, after including response time data, the recovered parameters achieve a smaller mean squared error (MSE, see online Supplementary Simulation recovery for detailed simulation procedure and results). Therefore, in the results part of this paper, we choose to report the estimation results from the two models as robustness checks.

For the comparison of the learning rates across conditions, we implement 2 (Group: HC *v.* MDD) × 2 (Valence: PE+, PE−) × 3 (Probability: Symmetric, Asymmetric, Reversal) repeated ANOVA for the partial feedback and 2 (Group: HC *v.* MDD) × 2 (Valence: PE+, PE−) × 2 (Selection: chosen *v.* unchosen) × 3 (Probability: Symmetric, Asymmetric, Reversal) for the complete feedback. Collapsing the 3 probabilities (Symmetric, Asymmetric, Reversal), we further did the 2 (Group: HC *v.* MDD) × 2 (Valence: PE+, PE−) × 2 (Selection: chosen *v.* unchosen) ANOVA for partial feedback and 2 (Group: HC *v.* MDD) × 2 (Valence: PE+, PE−) repeated measure ANOVA for complete feedback. Additionally, *post hoc* pairwise comparisons were implemented when the interaction terms were significant. As noted in the section 'Reinforcement learning model with standard random utility specification', we reported the general ANOVA results and the asymmetric condition in the Results section below and the other details in online supplementary material (see online Supplementary Computational model).

To formulate the connection between the severity of clinical symptoms and the learning behavior from the two-arm bandit task, we derived the learning index from the learning rate (see online Supplementary Learning index for detailed definition) and run the regression analysis with the HAMD score as the dependent variable and the learning index (partial, complete) as the independent variable using robust standard error.

All the data manipulation and statistical analysis including the computational modeling of the reinforcement model were implemented using the open-source software R (https://www.r-project.org, version 4.1.2). R package *lfe* (Gaure, 2013) was used for the linear regression including that adjusted for the clustered standard errors (payoff, *Box-Cox* transformed response time); package *rms* (Harrell, 2021) was adopted to run the multi-level logistic regression (choice accuracy); and *bruceR* (Bao, 2022) was used to perform the repeated-measure ANOVA analysis (two versions of RL model-derived learning rates). Two-sided *t* test was used for the statistical reports. Multiple comparisons were corrected using Bonferroni method when appropriate.

## Result

### Choice accuracy

For the monetary payoff, the multi-level regression analysis revealed that the healthy adolescent controls earned more token points (mean = 44.10, S.E. = 3.022) from the RL task than that of the MDD patients (mean = 35.29, S.E. = 2.90, online Supplementary Fig. S1, $\beta_{Group} = 0.031$, $p = 0.034$) (see online Supplementary Table S2). Consistent with our general intuition, there was also a prominent effect of information ($\beta_{Information} = -0.049$, $p < 0.001$). Both for MDD and HC group, the subjects had better performance with complete feedback compared to partial feedback (HC: $\beta_{Information} = -0.055$, $p < 0.001$; MDD: $\beta_{Information} = -0.043$, $p = 0.003$).

For the choice accuracy, the mixed-level logistic regression analysis reveals a similar finding as those observed for the payoff (see online Supplementary Table S3). In general, the MDD patient had worse performance over the option selection as opposed to the control group ($\beta_{Group} = 0.182$, $p = 0.042$). There was also a prominent effect for information ($\beta_{Information} = -0.306$, $p < 0.001$) such that the subjects showed better performance in the complete condition notwithstanding group (HC: $\beta_{Information} = -0.383$, $p < 0.001$; MDD: $\beta_{Information} = -0.235$, $p = 0.005$), and generally in line with the results of monetary payoff. For the pooled data with asymmetric condition and the first half of the reversal condition, we found a prominent group effect ($\beta_{Group} = 0.360$, $p = 0.016$), and the interaction between group and information was also marginally significant ($\beta_{Group\times Information} = -0.259$, $p = 0.087$).

Additionally, collapsed by condition (asymmetric and reversal) over choice accuracy, irrespective of whether it is the asymmetric or the 1st half of the reversal condition (see online Supplementary Table S3), we found that there was a prominent effect of group (asymmetric: $\beta_{Group} = 0.354$, $p = 0.036$; reversal 1st half: $\beta_{Group} = 0.381$, $p = 0.030$). Generally, the depressive subjects tended to perform worse than that of the healthy controls. As illustrated in Fig. 2, when checking the results for the partial and complete condition separately, we found that, regardless of the asymmetric or the reversal condition (see online Supplementary Table S4), for the partial condition, we failed to find a significant effect (asymmetric: $\beta_{Group} = 0.130$, $p = 0.391$; reversed 1st half: $\beta_{Group} = 0.038$, $p = 0.830$). However, the group effect was significant for the complete condition (asymmetric: $\beta_{Group} = 0.359$, $p = 0.037$; reversal 1st half: $\beta_{Group} = 0.388$, $p = 0.030$). Moreover, for the 2nd half of the reversal condition, no matter for the aggregate or those separated by information, we did not observe any significant difference (aggregate: $\beta_{Group} = 0.061$, $p = 0.632$; partial: $\beta_{Group} = 0.174$, $p = 0.216$; complete: $\beta_{Group} = 0.064$, $p = 0.632$).

### Response time

Considering the response time at the stage of choice execution, as presented in Fig. 3, we compared the difference of the *Box-Cox* transformed RT data between MDD and HC. The multi-level linear regression of the transformed response time reveals that, albeit the poor performance (lower payoff and choice accuracy) of the MDD subjects (see online Supplementary Table S5), they nevertheless exhibit a prolonged response time ($\beta_{Group} = -0.165$, $p < 0.001$). However, there is no effect for *information* ($\beta_{Information} = -0.014$, $p = 0.214$). As indicated in online Supplementary Table S5, the response time of the asymmetric and reversal conditions was longer than that of the symmetric condition (Asymmetric: $\beta_{Probability} = 0.016$, $p = 0.153$; Reversal: $\beta_{Probability} = 0.025$, $p = 0.058$), and the response time of the reversal was longer than that of the asymmetric condition ($\beta_{Probability} = 0.009$, $p = 0.439$). Therefore, we further examined the response time separately for symmetric, asymmetric and reversal condition.

Firstly, for the symmetric condition, the regression analysis showed a significant effect of group ($\beta_{Group} = -0.175$, $p < 0.001$), but no effect for information ($\beta_{Information} = -0.025$, $p = 0.234$). Further analysis suggests that irrespective of whether the partial or complete (see online Supplementary Table S6), the response

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

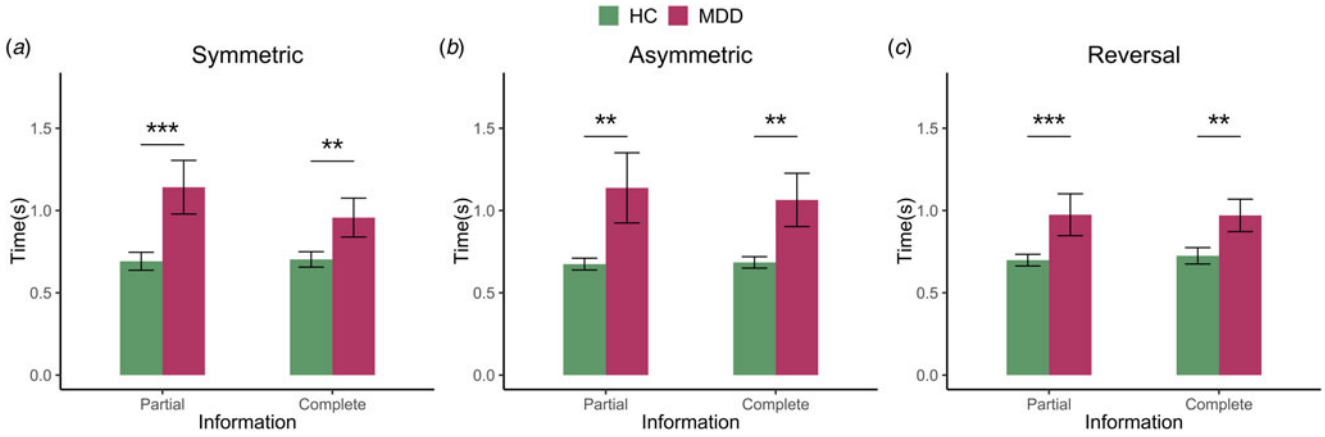


**Figure 3.** This figure exhibits the response time for MDD and HC across three conditions, i.e. symmetric, asymmetric and reversal condition with partial and complete information stratification. Error bars describe standard errors (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

times of MDD were longer than that HC (partial: $\beta_{Group} = -0.208$, $p < 0.001$; complete: $\beta_{Group} = -0.143$, $p = 0.008$). Secondly, for the asymmetric condition, the regression analysis also showed a significant effect of group ($\beta_{Group} = -0.156$, $p = 0.001$), but there was no effect for information condition ($\beta_{Information} = -0.029$, $p = 0.121$). The separate analysis indicated a similar effect as those observed for the symmetric condition, viz. the response time of HC was shorter than MDD for partial condition ($\beta_{Group} = -0.162$, $p = 0.002$) and complete condition ($\beta_{Group} = -0.150$, $p = 0.004$). Finally, for the reversal condition, the regression analysis showed a significant effect of group ($\beta_{Group} = -0.165$, $p < 0.001$), but no effect for information ($\beta_{Information} = 0.012$, $p = 0.487$). The partial and complete differentiation also indicated RT difference between HC and MDD (partial: $\beta_{Group} = -0.172$, $p < 0.001$; complete: $\beta_{Group} = -0.158$, $p = 0.003$). Therefore, for the RT comparison, we find that for all three conditions across symmetric, asymmetric and reversal condition, there was a stable and prominent difference across MDD and healthy subjects, indicating the important role of the reaction time in the process of decision making for the learning task.

### Computational modelling

For the computational modeling estimation, we applied the adapted RW reinforcement learning framework as illustrated in equation 1 and 3 (Niv, Edlund, Dayan, & O'Doherty, 2012; Palminteri et al., 2017). Such a framework allows us to test both the potential asymmetric/symmetric coding of the positive and negative RPE and their pattern in the chosen and unchosen condition.

On the basis of standard reinforcement learning model (Model I), given the prominent RT discrepancy between the two groups (Fig. 3) and potential better unbiased specification, according to the suggestion of Webb (2019), we further introduced the response time into the RL framework, and examine the learning rate for normal adolescent subjects and MDD patients accordingly (Model II). As revealed in Fig. 4, the general pattern was similar to what was observed in the RL model (see online Supplementary Fig. S2), but with a potentially larger contrast between the normal subjects and MDD patients. A possible conundrum in our results is that, since we separately estimate the models with the data from each condition, the observations in the asymmetric condition might be imbalanced among the four parameters. If the observations of PEs are very few for some parameters, the estimation result could be highly biased. To rule out this possibility, based on the results of Model I, we have counted the realized observations of PEs corresponding to each parameter for each individual and reported the average distribution in each condition across all individuals (see online Supplementary Fig. S5). The results show that, in the asymmetric condition,

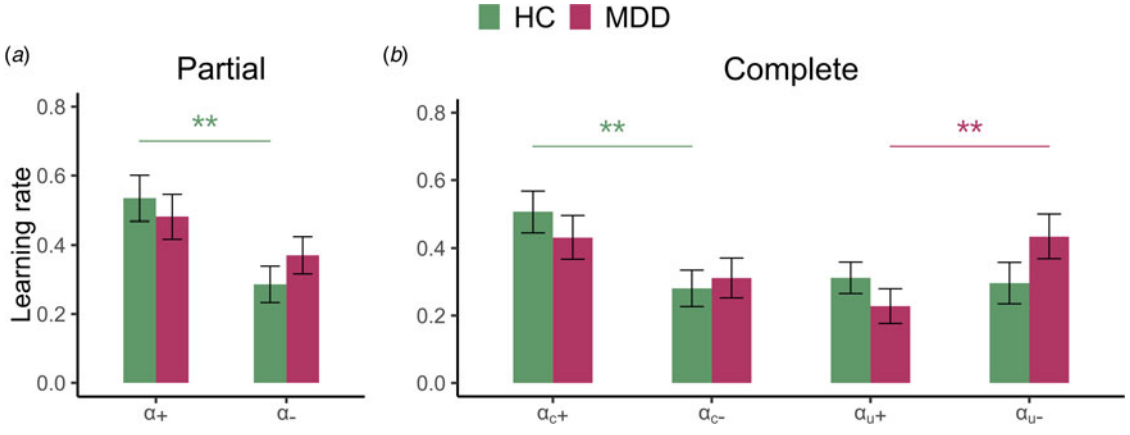


**Figure 4.** The learning rates of RL model with bounded accumulation from the asymmetric condition. The green color is for the healthy controls and the red color is for the MDD patients. Error bars describe standard errors (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

the distribution of the observations is not especially biased and there is no significant imbalance in the data for the estimation exercise.

The 3 × 2 × 2 repeated ANOVA in the partial condition, it showed a significant effect of probability ($F(2, 164) = 5.667$, $p = 0.004$, $\eta^2 p = 0.065$). Although there was no effect for group ($F(1, 82) = 0.288$, $p = 0.593$, $\eta^2 p = 0.004$), there was a prominent interaction effect between probability and valence ($F(2, 164) = 3.844$, $p = 0.023$, $\eta^2 p = 0.045$). For the 3 × 2 × 2 × 2 repeated ANOVA in the complete condition, a significant effect of probability was observed ($F(2, 164) = 5.667$, $p = 0.004$, $\eta^2 p = 0.065$), and a significant effect of selection ($F(1, 82) = 4.500$, $p = 0.037$, $\eta^2 p = 0.052$). There was no effect for group ($F(1, 82) = 1.249$, $p = 0.267$, $\eta^2 p = 0.015$). There was also a prominent effect between probability and selection ($F(2, 164) = 4.081$, $p = 0.019$, $\eta^2 p = 0.047$). Hence, there were different learning rates under different probability conditions, especially in the complete condition. Moreover, the selection condition also affected the learning rate. Here we mainly focused on the asymmetric condition, and the remainder of the results is detailed in the online Supplementary material (see online Supplementary Computational model, Fig. S3-S4).

As illustrated in Fig. 4, for the healthy adolescent subjects, there was a positive bias in the partial condition, viz. there was a higher learning rate for PE+ than that of PE− ($\alpha_+ = 0.535$, S.E. = 0.065, $\alpha_- = 0.286$, S.E. = 0.053, $t(82) = 2.974$, $p = 0.004$). For the complete condition, there was also a positive bias for the chosen option ($\alpha_{c+} = 0.506$, S.E. = 0.063, $\alpha_{c-} = 0.281$, S.E. = 0.057, $t(82) = 2.732$, $p = 0.008$) and there was no discrepancy for the unchosen option ($\alpha_{u+} = 0.312$, S.E. = 0.049, $\alpha_{u-} = 0.296$, S.E. = 0.063, $t(82) = 0.205$, $p = 0.838$). However, with respect to the adolescent MDD group, however, we failed to find the learning rate difference between positive negative RPE irrespective of whether it is the chosen option in the partial condition ($\alpha_+ = 0.481$, S.E. = 0.065, $\alpha_- = 0.370$, S.E. = 0.053, $t(82) = 1.326$, $p = 0.189$) or the chosen option in the complete condition ($\alpha_{c+} = 0.431$, S.E. = 0.063, $\alpha_{c-} = 0.312$, S.E. = 0.057, $t(82) = 1.450$, $p = 0.151$). Strikingly, for the RT, there remained a prominent pattern of the non-negligible negative bias for the unchosen option and a higher learning rate for PE- than that of PE+ ($\alpha_{u+} = 0.228$, S.E. = 0.049, $\alpha_{u-} = 0.434$, S.E. = 0.063, $t(82) = -2.713$, $p = 0.008$).

Finally, to examine the feasibility of verifying whether the parameters derived from the computational model indeed reflect the severity of the clinical symptom of the depressive disease, we ran an individual heterogeneity analysis for the adolescent MDD patients and considered whether the constructed learning index from partial and complete condition could link with the scores in the HAMD questionnaire. We found that there was a significantly negative connection between learning index and HAMD score ($\beta_{LI} = -2.402$, $p = 0.025$) (see Fig. 5 and online Supplementary Table S9). For the partial condition, the result was $\beta_{LI} = -2.926$, $p = 0.078$, and $\beta_{LI} = -2.197$, $p = 0.074$ for the complete condition (see online Supplementary Table S10).

## Discussion

The current study applies an instrumental learning task with the manipulation of the degree of the revealed information to evaluate the reinforcement learning behavior for adolescent depressive patients. Both the earned payoff and the choice accuracy reveal that the depressive adolescent patients generally perform poorly, compared with normal age-matched controls. As previously

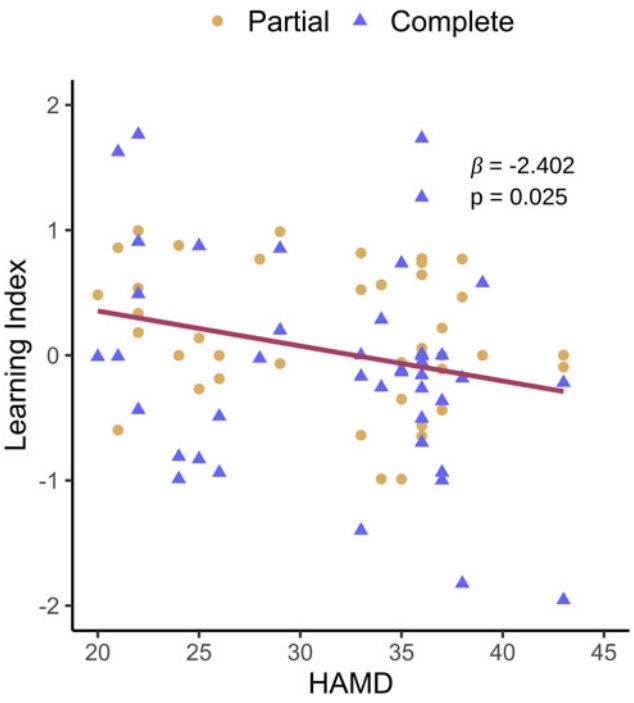


**Figure 5.** It illustrates the links between the HAMD depression score and the learning index for the partial and complete condition.

reported, the patients also displayed longer reaction times to complete the choice selection (Bakic et al., 2017; Chase et al., 2010; Pizzagalli, Iosifescu, Hallett, Ratner, & Fava, 2008). Computational modeling reveals that the usually observed positive bias in typical adolescents is absent in the depressive subjects. Notably, as compared with the normal subjects, the depressive subjects also exhibit an increased bias toward choosing the negative outcome for the counterfactual option. Finally, the learning index derived from the bias of the learning rate is correlated with the severity of the depressive symptoms of adolescent patients.

The behavioral findings that depressed adolescents choose negative outcomes, as well as the results from the computational modeling where we observe an increased negative prediction error for the unchosen option, viz. compared with the positive RPE there is a higher learning rate for the negative RPE in the depressed adolescent group, allows us to critically evaluate recent advances regarding how depressed adolescent subjects relate to positive and negative RPE, i.e. whether in a symmetric or asymmetric manner (Fig. 4).

Consistent with recent advances obtained in empirical RL studies (Lefebvre et al., 2017; Niv et al., 2012; Sugawara & Katahira, 2021), for the chosen option no matter in the partial or complete condition, normal adolescent subjects tend to exhibit a higher learning rate toward the positive RPE compared to negative RPE (Fig. 4). On the one side, these results underscore that healthy individuals are characterized by an optimistic bias and who by and large, tend to update the belief that matches self-serving interests or opinions (Kappes, Harvey, Lohrenz, Montague, & Sharot, 2019; Sharot & Garrett, 2016). Furthermore, such a tendency is also reliably observed in the dynamic instrumental learning context (Lefebvre et al., 2017). For the counterfactual outcome (unchosen option), the normal adolescent subjects do not show this tendency. That being said, some recent studies suggest perhaps there is an inclination to

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

such a confirmation bias even in non-clinical subjects. Adult subjects tend to have an opposite learning rate pattern for the counterfactual outcome (Palminteri et al., 2017; Tarantola et al., 2021). Nevertheless, at least one recent study suggests that, compared with adults, the typical adolescent subjects seem to at least partially fail to consider the alternative information from the unchosen option (Palminteri et al., 2016).

Focusing on the adolescent depressive subjects, we find that, similar to adult depressive patients who tend to reduce the response toward positive rewards and are more sensitive to negative stimuli, adolescent depressed patients, similar to their adult counterparts, are characterized by an omitted positive bias (Auerbach, Pagliaccio, & Pizzagalli, 2019; Bishop & Gagne, 2018; Kube et al., 2020; Nielson et al., 2021). Therefore, in contrast to the normal subjects who disregard the negative RPE, the depressive subjects apparently encode the prediction error with positive and negative valence in a more symmetric manner, consistent with the notion that the depressive patients in fact tend to objectively interpret the choice-contingent outcome from a more 'realistic' manner (depressive realism, Frank, 2016; Seidel et al., 2012).

Interestingly, for the depressive adolescent patients, with respect to the learning rate, they have a pronounced negative bias toward the outcome of the counterfactual option (Fig. 4). Seemingly counterintuitive at first glance, this is nevertheless, in accordance with the findings that the depressive subjects show more regret toward the post decision outcome (Kraines, Krug, & Wells, 2017; Roese et al., 2009). As we noted above, whether depressive patients show a negative bias or not is an unresolved question that is recently experiencing a heated debate (Brolsma et al., 2021). Importantly, the findings from the present study offer a potentially new perspective toward understanding more precisely the exact role of negative bias in depression. Notably, when the choice and RPE valence jointly come into play, it is not only possible to check whether there is a reduced positive bias or increased negative bias, but also feasible to test the asymmetry of the chosen *v.* unchosen PE (e.g. Palminteri et al., 2016). We infer that, given the general self-blame inclination for the justification of choice selection observed in the depressive subjects, such atypical subjects have a predisposition to exhibit more counterfactual thinking. Hence, they show a higher response toward counterfactual outcome and tend to be more responsive toward the unchosen option, which leads to an increased pessimistic bias toward the unchosen outcome (Broomhall, Phillips, Hine, & Loi, 2017).

With respect to the adolescent depressive subjects, the degree of bias at the individual level from the factual and counterfactual outcomes, are correlated with the severity of the depressive symptoms as measured by the HAMD depression score (Fig. 5). This finding suggests that the anomalous behavioral bias toward the RPE is a possibly salient mechanistic channel that underpins the decision deficits which is prominent in adolescent depressive orders. Therefore, by leveraging the instrumental learning task with computational modeling, and including the model with response time which likely reflects the process of dynamics of intra-trial process, the current study suggests there is considerable value to search for computational and decision markers in depression. Finding such unique biomarkers, would undoubtedly not only enhance our theoretical understanding of depression but also contribute to improvements in clinical evaluation such as the therapeutic effect of drugs and cognitive behavioral therapy. Further studies could profitably integrate the instrumental learning task with neuroimaging techniques (e.g. fMRI) to directly validate the findings we observed here in order to examine the extent to which both choice and RPE valence shapes the behavioral regularities as well as observed deficits in depressed adolescent subjects.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/S0033291723001307.

**Financial support.** This work was supported by grant 71971199 and 71942004 from the National Natural Science Foundation of China.

**Conflict of interest.** None.

## References

Auerbach, R. P., Pagliaccio, D., & Pizzagalli, D. A. (2019). Toward an improved understanding of anhedonia. *JAMA Psychiatry*, *76*(6), 571–573. doi:10.1001/jamapsychiatry.2018.4600.

Bakic, J., Pourtois, G., Jepma, M., Duprat, R., De Raedt, R., & Baeken, C. (2017). Spared internal but impaired external reward prediction error signals in major depressive disorder during reinforcement learning. *Depression and Anxiety*, *34*(1), 89–96. doi:10.1002/da.22576.

Bao, H. W. S. (2022). bruceR: Broadly useful convenien and efficient R functions. R package version 0.8.x. Retrieved from https://CRAN.R-project.org/package=bruceR.

Bavard, S., Rustichini, A., & Palminteri, S. (2021). Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Science Advances*, *7*(14), eabe0340. doi:10.1126/sciadv.abe0340.

Berwian, I. M., Wenzel, J. G., Collins, A. G., Seifritz, E., Stephan, K. E., Walter, H., & Huys, Q. J. (2020). Computational mechanisms of effort and reward decisions in patients with depression and their association with relapse after antidepressant discontinuation. *JAMA Psychiatry*, *77*(5), 513–522. doi:10.1001/jamapsychiatry.2019.4971.

Bishop, S. J., & Gagne, C. (2018). Anxiety, depression, and decision making: A computational perspective. *Annual Review of Neuroscience*, *41*, 371–388. doi:10.1146/annurev-neuro-080317-062007.

Brolsma, S. C., Vassena, E., Vrijsen, J. N., Sescousse, G., Collard, R. M., van Eijndhoven, P. F., & …Cools, R. (2021). Negative learning bias in depression revisited: Enhanced neural response to surprising reward across psychiatric disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(3), 280–289. doi:10.1016/j.bpsc.2020.08.011.

Bromberg-Martin, E. S., & Sharot, T. (2020). The value of beliefs. *Neuron*, *106*(4), 561–565. doi:10.1016/j.neuron.2020.05.001.

Broomhall, A. G., Phillips, W. J., Hine, D. W., & Loi, N. M. (2017). Upward counterfactual thinking and depression: A meta-analysis. *Clinical Psychology Review*, *55*, 56–73. doi:10.1016/j.cpr.2017.04.010.

Chambon, V., Thero, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, *4*(10), 1067–1079. doi:10.1038/s41562-020-0919-5.

Chase, H. W., Frank, M. J., Michael, A., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2010). Approach and avoidance learning in patients with major depression and healthy controls: Relation to anhedonia. *Psychological Medicine*, *40*(3), 433–440. doi:10.1017/S0033291709990468.

Clayborne, Z. M., Varin, M., & Colman, I. (2019). Systematic review and meta-analysis: Adolescent depression and long-term psychosocial outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(1), 72–79. doi:10.1016/j.jaac.2018.07.896.

Fontanesi, L., Gluth, S., Spektor, M. S., & Rieskamp, J. (2019). A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review*, *26*(4), 1099–1121. doi:10.3758/s13423-018-1554-2.

Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science (New York, N.Y.)*, *306*(5703), 1940–1943. doi:10.1126/science.1102941.

Frank, R. H. (2016). *Success and luck*. Princeton: Princeton University Press.

Gaure, S. (2013). lfe: Linear group fixed effects. *The R Journal*, *5*(2), 104–116. doi:10.32614/RJ-2013-031.

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 523–536. doi:10.3758/s13415-015-0347-6.

Harrell, F. E. Jr. (2021). rms: Regression Modeling Strategies. R package version 6.2-0. Retrieved from https://CRAN.R-project.org/package=rms.

Hartzmark, S. M., Hirshman, S. D., & Imas, A. (2021). Ownership, learning, and beliefs. *The Quarterly Journal of Economics*, *136*(3), 1665–1717. doi:10.1093/qje/qjab010.

Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2019). Confirmation bias in the utilization of others' opinion strength. *Nature Neuroscience*, *23*(1), 130–137. doi:10.1038/s41593-019-0549-2.

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine*, *44*(3), 579–592. doi:10.1017/S0033291713001074.

Kraines, M. A., Krug, C. P., & Wells, T. T. (2017). Decision justification theory in depression: Regret and self-blame. *Cognitive Therapy and Research*, *41*(4), 556–561. doi:10.1007/s10608-017-9836-y.

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry*, *87*(5), 388–398. doi:10.1016/j.biopsych.2019.07.017.

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., … Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology*, *43*(7), 1581–1588. doi:10.1038/s41386-018-0032-x.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*(4), 1–9. doi:10.1038/s41562-017-0067.

Lefebvre, G., Summerfield, C., & Bogacz, R. (2021). A normative account of confirmation bias during reinforcement learning. *Neural Computation*, *34*(2), 1–31. doi:10.1162/neco_a_01455.

Lu, W. (2019). Adolescent depression: National trends, risk factors, and healthcare disparities. *American Journal of Health Behavior*, *43*(1), 181–194. doi:10.5993/AJHB.43.1.15.

Ma, Y., Li, S., Wang, C., Liu, Y., Li, W., Yan, X., … Han, S. (2016). Distinct oxytocin effects on belief updating in response to desirable and undesirable feedback. *Proceedings of the National Academy of Sciences*, *113*(33), 9256–9261. doi:10.1073/pnas.1604285113.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Miletić, S., Boag, R. J., Trutti, A. C., Stevenson, N., Forstmann, B. U., & Heathcote, A. (2021). A new model of decision processing in instrumental learning tasks. *Elife*, *10*, e63055. doi:10.7554/eLife.63055.

Miller, L., & Campo, J. V. (2021). Depression in adolescents. *New England Journal of Medicine*, *385*(5), 445–449. doi:10.1056/NEJMra2033475.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, *16*(1), 72–80. doi:10.1016/j.tics.2011.11.018.

Mukherjee, D., Filipowicz, A. L. S., Vo, K., Satterthwaite, T. D., & Kable, J. W. (2020). Reward and punishment reversal-learning in major depressive disorder. *Journal of Abnormal Psychology*, *129*(8), 810–823. doi:10.1037/abn0000641.

Mullen, K., Ardia, D., Gil, D. L., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, *40*(6), 1–26. doi:10.18637/jss.v040.i06.

Ng, T. H., Alloy, L. B., & Smith, D. V. (2019). Meta-analysis of reward processing in major depressive disorder reveals distinct abnormalities within the reward circuit. *Translational Psychiatry*, *9*(1), 293. doi:10.1038/s41398-019-0644-x.

Nielson, D. M., Keren, H., O'Callaghan, G., Jackson, S. M., Douka, I., Vidal-Ribas, P., … Stringaris, A. (2021). Great expectations: A critical review of and suggestions for the study of reward processing as a cause and predictor of depression. *Biological Psychiatry*, *89*(2), 134–143. doi:10.1016/j.biopsych.2020.06.012.

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience*, *32*(2), 551–562. doi:10.1523/JNEUROSCI.5498-10.2012.

Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S. J. (2016). The computational development of reinforcement learning during adolescence. *PLoS Computational Biology*, *12*(6), e1004953. doi:10.1371/journal.pcbi.1004953.

Palminteri, S., Lefebvre, G., Kilford, E. J., & Blakemore, S. J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLoS Computational Biology*, *13*(8), e1005684. doi:10.1371/journal.pcbi.1005684.

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, *9*(12), 947–957. doi:10.1038/nrn2513.

Pedersen, M. L., & Frank, M. J. (2020). Simultaneous hierarchical Bayesian parameter estimation for reinforcement learning and drift diffusion models: A tutorial and links to neural data. *Computational Brain & Behavior*, *3*(4), 458–471. doi:10.1007/s42113-020-00084-w.

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review*, *24*(4), 1234–1251. doi:10.3758/s13423-016-1199-y.

Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008). Reduced hedonic capacity in major depressive disorder: Evidence from a probabilistic reward task. *Journal of Psychiatric Research*, *43*(1), 76–87. doi:10.1016/j.jpsychires.2008.03.001.

Raio, C. M., Hartley, C. A., Orederu, T. A., Li, J., & Phelps, E. A. (2017). Stress attenuates the flexible updating of aversive value. *Proceedings of the National Academy of Sciences*, *114*(42), 11241–11246. doi:10.1073/pnas.1702565114.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales* (pp. 223–237). Oxford: Oxford Pyschologists Press.

Roese, N. J., Epstude, K. A. I., Fessel, F., Morrison, M., Smallman, R., Summerville, A., … Segerstrom, S. (2009). Repetitive regret, depression, and anxiety: Findings from a nationally representative survey. *Journal of Social and Clinical Psychology*, *28*(6), 671–688. doi:10.1521/jscp.2009.28.6.671.

Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., … Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, *398*(10312), 1700–1712. doi:10.1016/s0140-6736(21)02143-7.

Seidel, E. M., Satterthwaite, T. D., Eickhoff, S. B., Schneider, F., Gur, R. C., Wolf, D. H., … Derntl, B. (2012). Neural correlates of depressive realism—An fMRI study on causal attribution in depression. *Journal of Affective Disorders*, *138*(3), 268–276. doi:10.1016/j.jad.2012.01.041.

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. doi:10.1016/j.cub.2011.10.030.

Sharot, T., & Garrett, N. (2016). Forming beliefs: Why valence matters. *Trends in Cognitive Sciences*, *20*(1), 25–33. doi:10.1016/j.tics.2015.11.002.

Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature*, *450*(7166), 102–105. doi:10.1016/j.cub.2011.10.030.

Sharot, T., Velasquez, C. M., & Dolan, R. J. (2010). Do decisions shape preference? Evidence from blind choice. *Psychological Science*, *21*(9), 1231–1235. doi:10.1177/0956797610379235.

Stevanovic, D., Jancic, J., & Lakic, A. (2011). The impact of depression and anxiety disorder symptoms on the health-related quality of life of children and adolescents with epilepsy. *Epilepsia*, *52*(8), e75–e78. doi:10.1111/j.1528-1167.2011.03133.x.

Stringaris, A., Vidal-Ribas Belil, P., Artiges, E., Lemaitre, H., Gollier-Briant, F., & Wolke, S., … IMAGEN Consortium. (2015). The brain's response to reward anticipation and depression in adolescence: Dimensionality, specificity, and longitudinal predictions in a community-based sample. *American Journal Psychiatry*, *172*(12), 1215–1223. doi:10.1176/appi.ajp.2015.14101298.

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press

Sugawara, M., & Katahira, K. (2021). Dissociation between asymmetric value updating and perseverance in human reinforcement learning. *Scientific Reports*, *11*(1), 3574. doi:10.1038/s41598-020-80593-7.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. Boston: MIT press.

Tarantola, T. O., Folke, T., Boldt, A., Perez, O. D., & De Martino, B. (2021). Confirmation bias optimizes reward learning. *bioRxiv*, *2021-02*. doi:10.1101/2021.02.27.433214.

Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E., & Binau, S. G. (2019). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology*, *128*(3), 185–199. doi:10.1287/mnsc.2017.2931.

Webb, R. (2019). The (neural) dynamics of stochastic choice. *Management Science*, *65*(1), 230–255. doi:10.1287/mnsc.2017.2931.

Wiehler, A., Chakroun, K., & Peters, J. (2021). Attenuated directed exploration during reinforcement learning in gambling disorder. *Journal of Neuroscience*, *41*(11), 2512–2522. doi:10.1523/JNEUROSCI.1607-20.2021.

Zhang, M., & He, Y. (2015). *Psychiatric rating scale manual*. Changsha: Hunan Science and Technology Press.

https://doi.org/10.1017/S0033291723001307 Published online by Cambridge University Press