AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Abnormal Psychology

ISSN: 0021-843X

2020, Vol. 129, No. 8, 810–823
http://dx.doi.org/10.1037/abn0000641

Reward and Punishment Reversal-Learning in Major Depressive Disorder

Dahlia Mukherjee
Penn State College of Medicine and The University of Pennsylvania

Alexandre L. S. Filipowicz
The University of Pennsylvania

Khoi Vo
Duke University

Theodore D. Satterthwaite and Joseph W. Kable
The University of Pennsylvania

Depression has been associated with impaired reward and punishment processing, but the specific nature of these deficits is still widely debated. We analyzed reinforcement-based decision making in individuals with major depressive disorder (MDD) to identify the specific decision mechanisms contributing to poorer performance. Individuals with MDD (n = 64) and matched healthy controls (n = 64) performed a probabilistic reversal-learning task in which they used feedback to identify which of two stimuli had the highest probability of reward (reward condition) or lowest probability of punishment (punishment condition). Learning differences were characterized using a hierarchical Bayesian reinforcement learning model. Depressed individuals made fewer optimal choices and adjusted more slowly to reversals in both the reward and punishment conditions. Computational modeling revealed that depressed individuals showed lower learning-rates and, to a lesser extent, lower value sensitivity in both the reward and punishment conditions. Learning-rates also predicted depression more accurately than simple performance metrics. These results demonstrate that depression is characterized by a hyposensitivity to positive outcomes, but not a hypersensitivity to negative outcomes. Additionally, we demonstrate that computational modeling provides a more precise characterization of the dynamics contributing to these learning deficits, offering stronger insights into the mechanistic processes affected by depression.

General Scientific Summary
A key symptom of depression is a decreased motivation to seek out positive experiences. This study finds that patients with major depressive disorder (MDD) have difficulty learning from positive outcomes and, to a lesser extent, also value positive outcomes to a lesser degree. This provides important insights into the specific mechanisms that could be impaired by major depressive episodes, and potentially targeted through both pharmacological and psychological interventions.

Keywords: major depressive disorder, reinforcement learning, reversal-learning, decision-making, computational psychiatry

Supplemental materials: http://dx.doi.org/10.1037/abn0000641.supp

Major depressive disorder (MDD) is often characterized by abnormal responses to affective stimuli, such as rewards and punishments (Chen, Takahashi, Nakagawa, Inoue, & Kusumi, 2015; Eshel & Roiser, 2010). Although there is widespread

This article was published Online First October 1, 2020.

Dahlia Mukherjee, Department of Psychiatry, Penn State College of Medicine, and Department of Psychology, The University of Pennsylvania; Alexandre L. S. Filipowicz, Departments of Psychology and Neuroscience, The University of Pennsylvania; Khoi Vo, Department of Psychology and Neuroscience, Duke University; Theodore D. Satterthwaite, Department of Psychiatry, The University of Pennsylvania; Joseph W. Kable, Department of Psychology, The University of Pennsylvania.

This research was supported by the Penn Psychology Graduate Fund (Dahlia Mukherjee), an NIMH F32 postdoctoral fellowship (Alexandre L. S. Filipowicz), and by NIH Grants R01 DA029149, R01 MH098899 and R01 CA170297 (Joseph W. Kable).

The data used in this article has been published previously (doi:10.1038/s41598-020-60230-z). Early versions of this article have been posted as a preprint (doi:10.31234/osf.io/aqgx3), and presented at Society for Neuroeconomics, 2019.

This study was approved by the Institutional Review Board at the University of Pennsylvania (IRB Project: 815189).

Dahlia Mukherjee and Alexandre L. S. Filipowicz are co–first authors.

Correspondence concerning this article should be addressed to Dahlia Mukherjee, who is now at the Department of Psychiatry and Behavioral Health, Penn State College of Medicine, Penn State Milton S. Hershey Medical Center, H073, 500 University Drive, Rm C5613A, Hershey, PA 17033. E-mail: dmukherjee@pennstatehealth.psu.edu

evidence that individuals with MDD respond differently to rewards and punishments than healthy individuals, the mechanisms driving these differences are less clear (Chen et al., 2015).

One prominent idea is that depression is associated with hyposensitivity to reward and hypersensitivity to punishment (Eshel & Roiser, 2010). In the reward domain, one line of research has examined how perceptual judgments are biased by asymmetric rewards. Compared to nondepressed individuals, depressed individuals show lower levels of reward bias (Pizzagalli, Iosifescu, Hallett, Ratner, & Fava, 2008; Pizzagalli, Jahn, & O'Shea, 2005). This decreased reward bias persists during remission and predicts worse outcomes during treatment (Pechtel, Dutra, Goetz, & Pizzagalli, 2013). In contrast, depression has also been linked, albeit inconsistently, with hypersensitivity to errors in cognitive tasks (Beats, Sahakian, & Levy, 1996; Chen et al., 2015; Elliott et al., 1996; Eshel & Roiser, 2010). Depressed individuals sometimes overreact to errors they commit, which impairs their subsequent performance (Beats et al., 1996; Elliott et al., 1996; Eshel & Roiser, 2010; Steffens, Wagner, Levy, Horn, & Krishnan, 2001).

Sensitivity to positive and negative feedback has also been studied in instrumental learning tasks, where subjects learn to select favorable actions on the basis of probabilistic feedback. In these tasks, learning tendencies have been characterized by how often subjects repeat actions after positive feedback ("win-stay") or switch actions after negative feedback ("lose-shift"). Some studies have found that depressed individuals show more 'lose-shift' responses than healthy controls, suggesting a heightened sensitivity to negative feedback (Dombrovski et al., 2015; Dombrovski, Szanto, Clark, Reynolds, & Siegle, 2013; Murphy, Michael, Robbins, & Sahakian, 2003). However, findings have been mixed, with other studies failing to find depression-related differences in 'lose-shift' responses (Chase et al., 2010; Dombrovski et al., 2010; Gradin et al., 2011).

Here we directly compare how MDD affects learning in response to rewards and punishments under structurally identical task conditions. We used two incentivized probabilistic reversal-learning tasks where subjects had to learn either to maximize reward or minimize punishment. In another report, we found that depressed individuals make fewer optimal choices than controls on these tasks, and that these differences in probabilistic reversal-learning are larger than on several other decision-making tasks (Mukherjee, Lee, Kazinka, Satterthwaite, & Kable, 2020). Here we specifically examine whether similar mechanisms account for impaired learning from rewards versus punishments.

We address these questions using computational modeling, as it allows us to distinguish between different potential causes for impairment (Huys, Maia, & Frank, 2016; Montague, Dolan, Friston, & Dayan, 2012; Wang & Krystal, 2014). Reinforcement learning (RL) models characterize how feedback is used to learn (Sutton & Barto, 1998) and are commonly used to study reward processing in depressed individuals (Chen et al., 2015; Huys, Daw, & Dayan, 2015; Huys, Pizzagalli, Bogdan, & Dayan, 2013). RL models can dissociate two different sources of learning deficits: changes in learning-rate (how much subjects adjust their behavior in response to feedback) and changes in value sensitivity (differences in the subjective value of outcomes). In the reward domain, early modeling attempts have attributed depression-related learning deficits to lower learning-rates (Chase et al., 2010; Chen et al., 2015). However, more recent studies argue that, rather than affecting learning, depression lowers value sensitivity (Huys et al., 2013; Huys et al., 2015).

Here we used computational modeling to identify a) the mechanisms that account for the deficits observed in reward and punishments learning, b) whether these mechanisms differ between reward and punishment learning, and c) whether fit RL parameters provide a benefit over-and-above behavioral metrics in classifying individuals with MDD. Overall we find that similar mechanisms contributed to poor MDD performance in both reward and punishment learning, with depressed individuals showing reduced learning-rates and, to a lesser extent, reduced value sensitivity. We also find that learning-rates were the strongest predictor of depression, outperforming other model parameters and behavioral metrics. Together these findings provide more robust insights into the specific feedback processing mechanisms affected by depression.

Method and Materials

Participants

128 subjects (64 diagnosed with MDD and 64 healthy controls) were recruited for the study. MDD subjects were recruited through flyers in the Department of Psychiatry and Behavioral Health, Counseling and Psychological Services and the Hospital of the University of Pennsylvania, and through referrals from treatment studies in the Department of Psychiatry. Healthy controls were recruited through flyers posted in the Departments of Psychology and Psychiatry, the Law School, the Graduate Student Office, and the Hospital. Subjects likely to meet study criteria based on an initial phone screen were invited for a diagnostic interview. MDD subjects were enrolled if (a) they met the diagnostic criteria for current MDD episode, (b) had no history of substance abuse/dependence in the past 6 months, and (c) had no history of bipolar disorder and/or psychotic episodes. Potential comorbidities beyond bipolar, substance use or psychotic disorders were not assessed. Diagnostic criteria were determined based on the Structured Clinical Interview for *DSM–IV Axis I Disorders* (SCID-I; First, Spitzer, Gibbon, & Williams, 2002). Of the 64 MDD subjects, 36 were referred directly from other clinical studies or service. Within the MDD group, 43 were undergoing treatment for depression (therapy and/or medication) and 19 were not (treatment data missing for 2 individuals). Inclusion criteria for controls were absence of current/past psychiatric illness, as assessed by the SCID, and absence of any psychotropic medications (see Table 1 for MDD and control demographic information). MDD subjects were moderately to severely depressed, with significantly higher BDI-II scores than controls (Supplementary Table 2), but did not differ significantly from controls in gender, race, education, age or cognitive ability (see Table 1).

The study was approved by the Institutional Review Board at the University of Pennsylvania (IRB #815189) and all subjects provided written informed consent. The clinical interview and study procedures were conducted by a master's level trained clinical psychologist (DM). Data from the reward and punishment reversal-learning tasks were collected as part of a larger study investigating value-based decision-making in MDD (Mukherjee et al., 2020). Subjects were paid $15/hr and received an additional incentive based on their decisions in one randomly selected task

Table 1
MDD and Control Group Demographics

Demographic information	Depressed		Control		
	N	%	*N*	%	χ^2-test *p*-value
Gender					.376
Female	37	57.8	33	51.6	
Male	27	42.2	31	48.4	
Race					.756
Black or African-American	34	53.1	33	51.6	
White	24	37.5	23	35.9	
Asian	4	6.2	7	10.9	
Other	2	3.1	1	1.6	
Education					.313
No high school	4	6.3	1	1.6	
High school	26	41.3	23	35.9	
Associate's	10	15.9	12	18.8	
Bachelor's	12	19	16	25	
Master's	11	17.5	9	14.1	
Doctoral	0	0	3	4.7	
Treatment					
No treatment	19	29.7	63	98.4	
Medication only	10	15.6	0	0	
Therapy only	14	21.9	0	0	
Both	19	29.7	0	0	
Missing	2	3.1	1	1.6	

	M	*SD*	*M*	*SD*	*p*-value
Age (years)	40.45	13.48	38.53	11.73	.299
WASI	101.32	14.63	103.70	14.09	.361

Note. MDD = major depressive disorder; WASI = Wechsler Abbreviated Scale Intelligence.

out of the eight administered. The components of the study were administered in the following order: SCID, eight decision-making tasks (including the two reversal-learning tasks), two subtests of the Wechsler Abbreviated Scale Intelligence Second edition (WASI-II; Wechsler, 2011). and seven self-report measures.

Measures

Clinical measures. Inclusion and exclusion diagnostic criteria were based on the SCID. Depression severity was assessed with the Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996). Current medication and psychotherapy treatment were self-reported.

Self-report measures. In addition to the BDI-II, subjects also completed the Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988), Depression, Anxiety and Stress Scale (Lovibond & Lovibond, 1995), Snaith Hamilton Pleasure Scale (Snaith et al., 1995), Rosenberg Self-Esteem Scale (Rosenberg, 1965), Behavioral Inhibition and Behavioral Activation Scale (Carver & White, 1994), and Cognitive Behavioral Avoidance Scale (Ottenbreit & Dobson, 2004).

Cognitive measure. Cognitive ability was assessed with the WASI-II. The matrix reasoning and similarities subtests were administered to obtain a full-scale IQ score. We chose the similarities subtest rather than the verbal reasoning subtest as it is less culturally biased (Razani, Murcia, Tabares, & Wong, 2007).

Reward probabilistic reversal-learning task. Subjects chose between two distinct fractal stimuli (Figure 1a), which were positioned randomly at one of two screen locations (left and right of a central white dot). On each trial, subjects pressed a keyboard button to choose one of the fractals. When subject responses resulted in a reward, a picture of a quarter was displayed to indicate the money added to their total winnings; otherwise, a red dot was displayed indicating that no money had been gained. The fractals were rewarded probabilistically, with the richer/poorer fractal rewarded 70%/30% of the time. Subjects were informed that on each trial, one fractal had a higher likelihood of delivering a reward, and that this association would reverse periodically throughout the task (i.e., the rich fractal from the previous trial would now have the lower probability and vice versa), but were not explicitly told which fractal was richer, or when reversals would occur. Subjects completed 4 practice trials before doing a full, 90-trial run, with reversals occurring every 30 trials.

Punishment probabilistic reversal-learning task. This task was similar to the reward task except that the goal was now to avoid choosing the fractal leading to punishment (signified by a red cross overlaying a quarter; Figure 1b). The "richer"/"poorer" fractal was punished 30%/70% of the time. Subjects were informed that one fractal led to more losses than the other and that this would switch periodically. The subjects started with $22.50 and lost $0.25 each time they received punishment feedback. Like the reward task, subjects did 4 practice trials before completing the full 90-trials run.

Performance Analysis

Behavioral metrics. We analyzed three basic trial-by-trial performance measures: (a) Rich choices, when subjects chose the

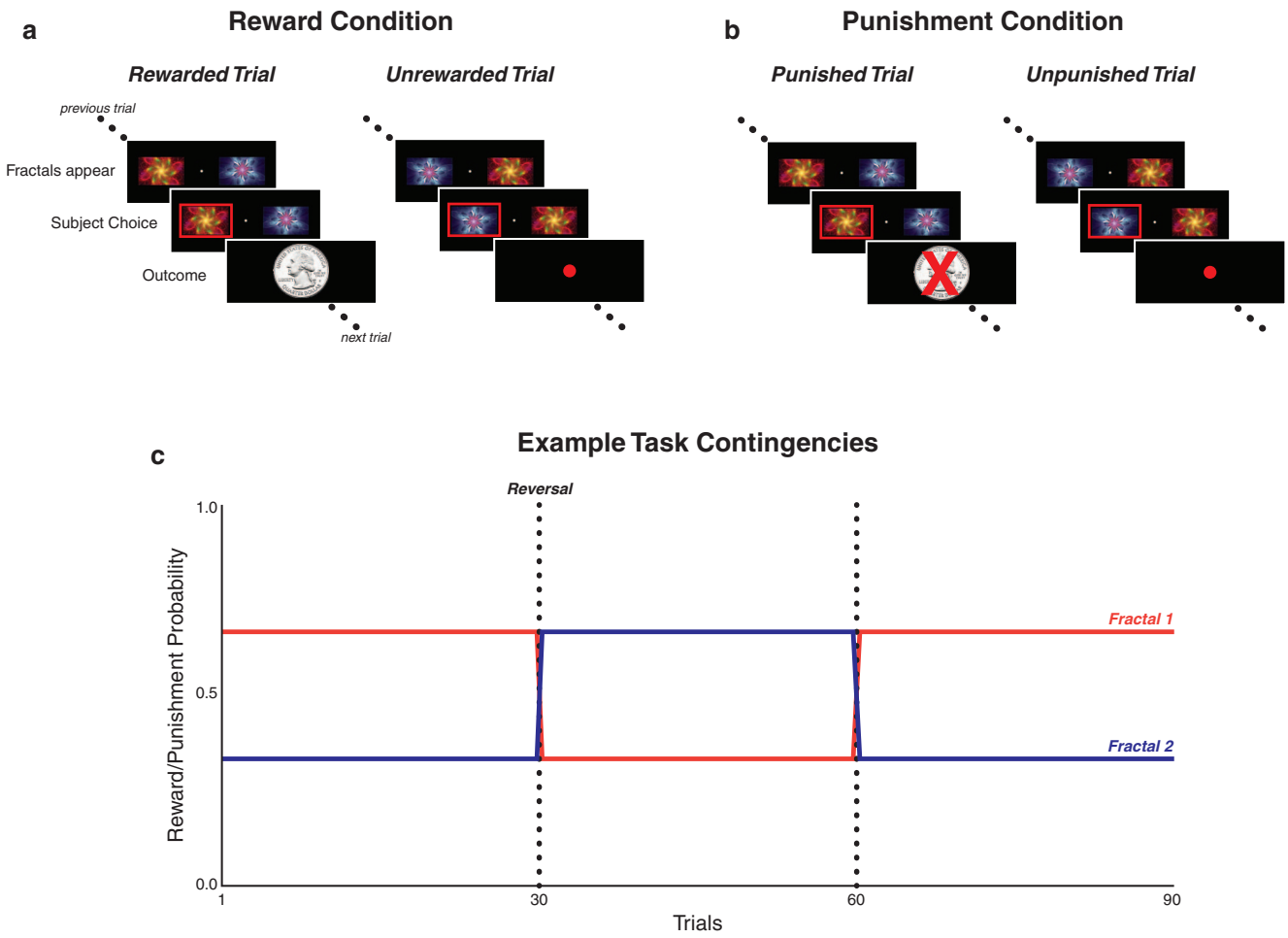

Figure 1. Reversal-learning task (a, b). At the start of each trial, subjects chose one of two fractals that they thought would either give them the highest chance of reward (reward condition) or lowest chance of punishment (punishment condition). (a) In the reward condition, a quarter appeared if the subject's choice resulted in a reward. (b) In the punishment condition, a quarter with a red X appeared if the subject's choice resulted in punishment. A red circle appeared for unrewarded or unpunished trials in the reward and punishment conditions respectively. (c) At the start of the task, one fractal had a 70% probability of reward/punishment whereas the other had a 30% probability. Reward/punishment probabilities switched every 30 trials. The fractal initially assigned the higher probability was counterbalanced across subjects. See the online article for the color version of this figure.

fractal with the higher positive outcome probability (i.e., reward/no punishment for the reward/punishment task), (b) Win-stay choices, when subjects chose the same fractal as the previous trial after a positive outcome, and (c) Lose-shift choices, when subjects switched fractal choices after a negative outcome (no reward/loss in the reward/punishment task).

Linear regression model. To test whether choice behavior was consistent with reinforcement learning, we fit a logistic regression to choice data (Lau & Glimcher, 2008) to estimate the influence of rewards (R) and choices (c) from previous trials (t) on the probability of choosing one fractal, $\Pr(c_t = f_1)$, versus the other, $\Pr(c_t = f_2)$ on the current trial:

$$\log\left(\frac{\Pr(c_t = f_1)}{\Pr(c_t = f_2)}\right) = \sum_{i=t-1}^{t-10} \beta_i(R_{f1,i} - R_{f2,i}) + \beta_{last\ choice}(c_{f1,t-1} - c_{f2,t-1}) + \beta_0 \quad (1)$$

Here β_i corresponds to the influence of reward received from up to 10 trials in the past on a subject's current fractal choice, $\beta_{last\ choice}$ captures the influence of a subject's last fractal choice (c_{f1} for fractal 1 and c_{f2} for fractal 2), and the β_0 intercept captures biases the subjects may have toward one of the fractals ($\beta_0 = 0$ indicates no bias).

Reinforcement learning models. Choice data from all subjects was fit using reinforcement learning models (Sutton & Barto, 1998). The models used the sequence of choices and outcomes to estimate the value (Q) of each fractal (f) for every trial. The fractal values are updated as trial-by-trial feedback is received using a standard update rule:

$$Q_{t+1}(f) = Q_t(f) + \alpha_f[R_t - Q_t(f)] \quad (2)$$

where R corresponds to the reward/outcome obtained after choosing the fractal f (coded 1/0 for reward/no reward in the reward task, and 1/0 for no punishment/punishment in the punishment task),

and α_f corresponds to a subject-specific learning-rate that governs the extent to which the observed feedback is used to update the value for the chosen fractal. The influence of differences in fractal values on decisions was captured by an inverse-temperature parameter β_f (see eq. 4 below), which could range between 0 (no influence of fractal values) and $+\infty$ (strong influence of fractal values). While traditionally associated with explore/exploit tradeoffs, this parameter is mathematically equivalent to other formulations of value sensitivity (Huys et al., 2013).

The model also included additional features that do not improve task performance but have been shown to influence subject performance on similar tasks (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Keung, Hagen, & Wilson, 2019). We first included a term to capture learning occurring regarding motor actions, rather than fractals. Values (Q) for actions (a) were tracked using a similar form to eq. 2.

$$Q_{t+1}(a) = Q_t(a) + [R_t - Q_t(a)] \quad (3)$$

For reasons highlighted in the model fitting section, we omitted a subject-specific learning-rate in eq. 3, effectively fixing the learning-rate for actions to a value of one. The influence of action values was captured by an action sensitivity parameter (β_a; eq. 4), which varied between -∞ and $+\infty$, as some subjects were *repelled* by the action value.

The model also included four parameters to capture response biases in subject choices. Two parameters captured perseverative tendencies toward repeating the same fractal choice (ρ_f) or same action (ρ_a) as the previous trial. Positive values indicate tendencies toward repeating the same fractal choice or motor action from the previous trial, regardless of outcome, whereas negative values correspond to tendencies toward switching choices or actions. Two final parameters captured tendencies toward choosing a particular fractal (π_f) or action (π_a) more than the other.

Together these parameters were used to compute the log odds, v, on any trial t that a subject would choose a fractal, with positive log odds corresponding to choosing fractal 1 (f_1), and the inverse logit corresponding to the probability that a subject would choose fractal 1 (eq. 5)

$$v_t = \beta_f[Q_t(f_1) - Q_t(f_2)] + \beta_a[Q_t(a = f_1) - Q_t(a = f_2)] + \rho_f + \rho_a + \pi_f + \pi_a \quad (4)$$

$$\Pr(f_1) = logit^{-1}(v_t) \quad (5)$$

The full model, with seven free parameters (α_f, β_f, β_a, ρ_f, ρ_a, π_{f1}, π_a), was fit to subjects' data separately for the reward and punishment conditions. To ensure that the full model provided the best fit for the data while accounting for model complexity, we compared fits from this full model to three RL models that used reduced forms of eq. 3–5: (1) simple RL (α_f, β_f), (2) RL plus action-learning (α_f, β_f, β_a), and (3) RL plus action-learning and perseveration (α_f, β_f, β_a, ρ_f, ρ_a). Model fits were compared using approximate leave-one-out cross-validation using the 'loo' R package (Vehtari, Gelman, & Gabry, 2017).

Model fitting and parameter comparisons. Models were fit as a multilevel hierarchical Bayesian model implemented in Stan (Carpenter et al., 2017) through the RStan interface (Stan Development Team, 2018). Such models provide more precise subject and group level parameters estimates (Huys et al., 2011; Huys et al., 2013). A subject i's parameters were estimated as joint multivariate distributions $\Pr(\mathbf{h}_i, \boldsymbol{\theta})$, with $\mathbf{h}$ corresponding to subject-specific parameter distributions (i.e., $\mathbf{h}_i \in \{\alpha_{f,i}, \beta_{f,i}, \ldots, \pi_{a,i}\}$). These multivariate parameter distributions were assumed to be generated from a group-level prior distribution $\Pr(\boldsymbol{\theta})$. Using this prior, we sought the posterior multivariate parameter distributions for each subject, given their data $\mathbf{D_i}$:

$$\Pr(\mathbf{h}_i, \boldsymbol{\theta}|\mathbf{D}_i) \propto \Pr(\mathbf{D}_i|\mathbf{h}_i, \boldsymbol{\theta})\Pr(\mathbf{h}_i|\boldsymbol{\theta})\Pr(\boldsymbol{\theta}) \quad (6)$$

Learning-rate parameters were assumed to be generated from a beta prior, fractal sensitivity parameters from a gamma prior, and all other parameters from Gaussian priors. We sampled from this posterior distribution using Stan's Hamiltonian MCMC algorithm. Six simultaneous chains were run with an 80,000 iteration warm-up (burn-in) period, and 170,000 posterior draws. Model convergence was assessed by ensuring that all $\hat{R}$ statistics were all near 1, $.99 < \hat{R} < 1.01$ (Gelman & Rubin, 1992), and effective sample sizes for each parameter were all greater than 1/100th the number of total samples. We additionally reduced our sampling step size (i.e., adapt $\delta = .99$) and implementing noncentered version of our models to eliminate sampling divergences (Betancourt & Girolami, 2015). Initial attempts to fit an action value learning-rate resulted in a high number of divergent samples, indicating that this parametrization was difficult to identify. These divergences were eliminated by assuming a fixed action learning-rate of 1. Individual subject parameters used for simulations and to predict each subject's depression status were generated using the mean values from each subject's parameter distributions (i.e., $E[\alpha_{f,i}]$, $E[\beta_{f,i}]$, . . . , $E[\pi_{a,i}]$).

Subject classifiers. To measure how well behavioral metrics and fit parameters predicted depression, we trained binary classifiers to predict out-of-sample subject depression status. Because behavioral metrics and parameters were correlated, we first performed a LASSO regression, tuned using leave-one-out cross-validation, to identify variables best able to classify depressed individuals. We used behavioral metrics and parameters with coefficients >0 in the LASSO regression in a subsequent binary classification analysis, and omitted those with coefficients = 0.

We next input the retained behavioral metrics and fit parameters into logistic classifiers trained to identify depressed individuals. These classifiers were trained using a repeated 10-fold cross-validation scheme (3 repetitions) using the "caret" package in R (Kuhn, 2008). The mean area under the receiver operating characteristic curve (AUC) for each repetition of each fold was used to assess out-of-sample classifier performance, where values of 1 indicate perfect classification accuracy and 0.5 indicate chance performance. Differences in model AUC values were compared using the Delong method (DeLong, DeLong, & Clarke-Pearson, 1988) implemented using the pROC package (Robin et al., 2011).

Additional statistics. Because the majority of our task data were non-normally distributed (e.g., proportion scores, beta-distributed parameter values), we used nonparametric signed-rank tests for pairwise comparisons and Spearman correlations for correlation analyses.

Results

Depressed Individuals Showed Poorer Reversal-Learning Than Healthy Controls

Depressed individuals performed worse than healthy controls on both reward and punishment reversal-learning. Depressed individuals made significantly fewer rich choices in both the reward (signed-rank test, $p = .033$) and the punishment conditions ($p = .003$; Figure 2).

We examined several aspects of performance to determine whether behavior in both reversal-learning tasks was consistent with reinforcement learning models. We first examined rich choices after a reversal. Consistent with reinforcement learning, both groups gradually adjusted their responses after a reversal, with the depressed group appearing to learn the new contingencies more slowly in both the reward and punishment conditions (Figure 2b).

Next we examined whether choices were more likely to be repeated after positive feedback, given that RL performance should show a high proportion of win-stay behavior. In both the reward and punishment conditions, each group had proportions of win-stay choices well above 50% (rank-sum test all ps $< 8.2 \times 10^{-11}$). However, the proportion of win-stay choices was significantly lower for the depressed group in both the reward (signed-rank test, $p = .005$) and punishment conditions ($p = .001$; Figure 3b). In contrast, the proportion of lose-shift responses did not differ between groups in either the reward ($p = .252$) or punishment conditions ($p = .165$; Figure 3). This suggests that while both groups show RL-like 'win-stay' responses, depressed subjects learned less from positive outcomes.

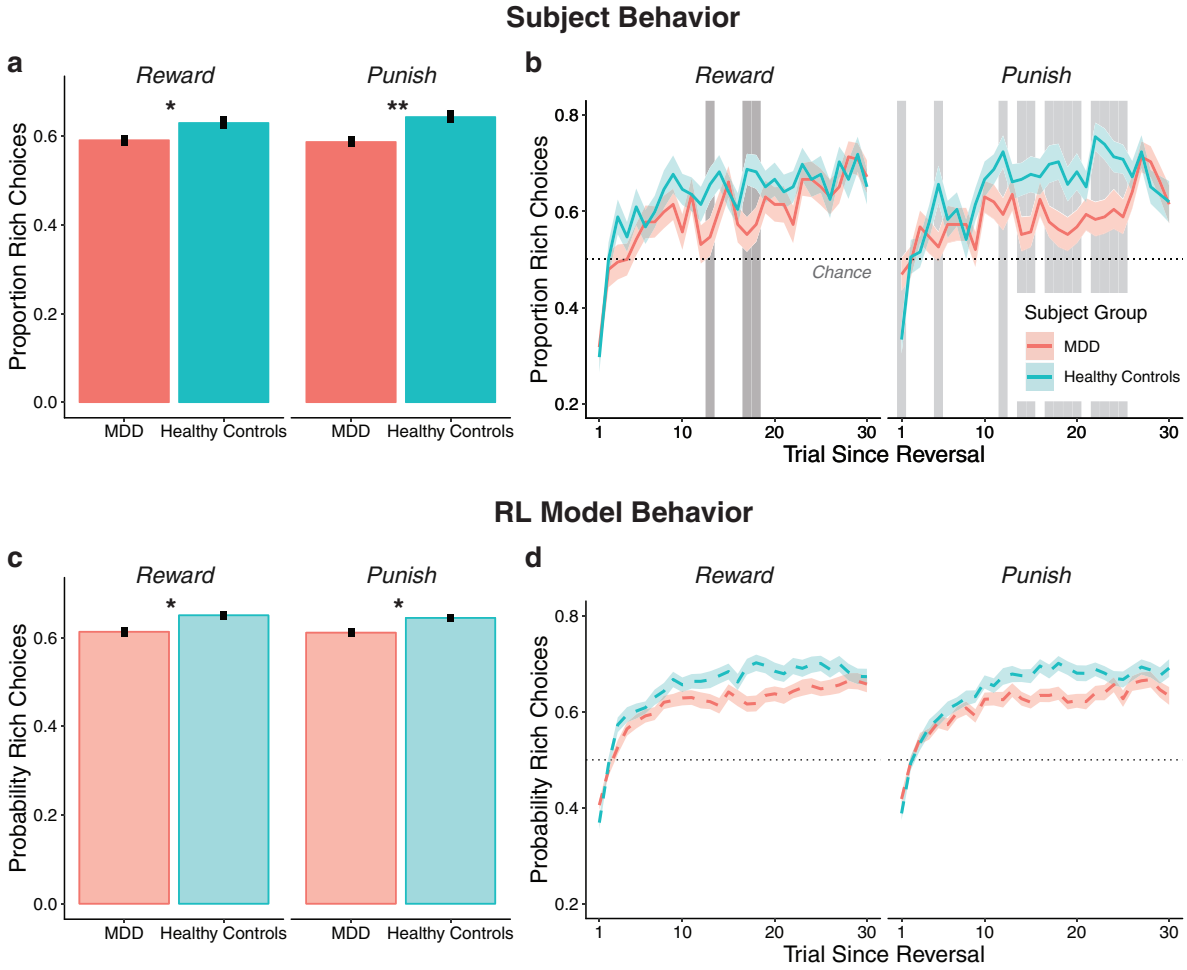

Figure 2. Individuals with MDD (red) made fewer rich choices than controls (blue) in both task conditions. (a) Mean proportion rich choices in MDD (red) and healthy controls (blue) across all trials in the reward and punishment task conditions. (b) Mean proportion rich responses as a function of trial since the last reversal (c–d). Identical statistics reported in (a) and (b) computed from simulations of subject behavior (100 independent simulations per subject using their fit parameter estimates). Bars and errorbars in (a, c) and lines and shading in (b, d) correspond to means and one standard error of the mean respectively. MDD = major depressive disorder; RL = reinforcement learning. Gray areas in (b) indicate trials after reversal on which signed-rank tests $p < .05$. * $p < .05$. ** $p < .01$. See the online article for the color version of this figure.

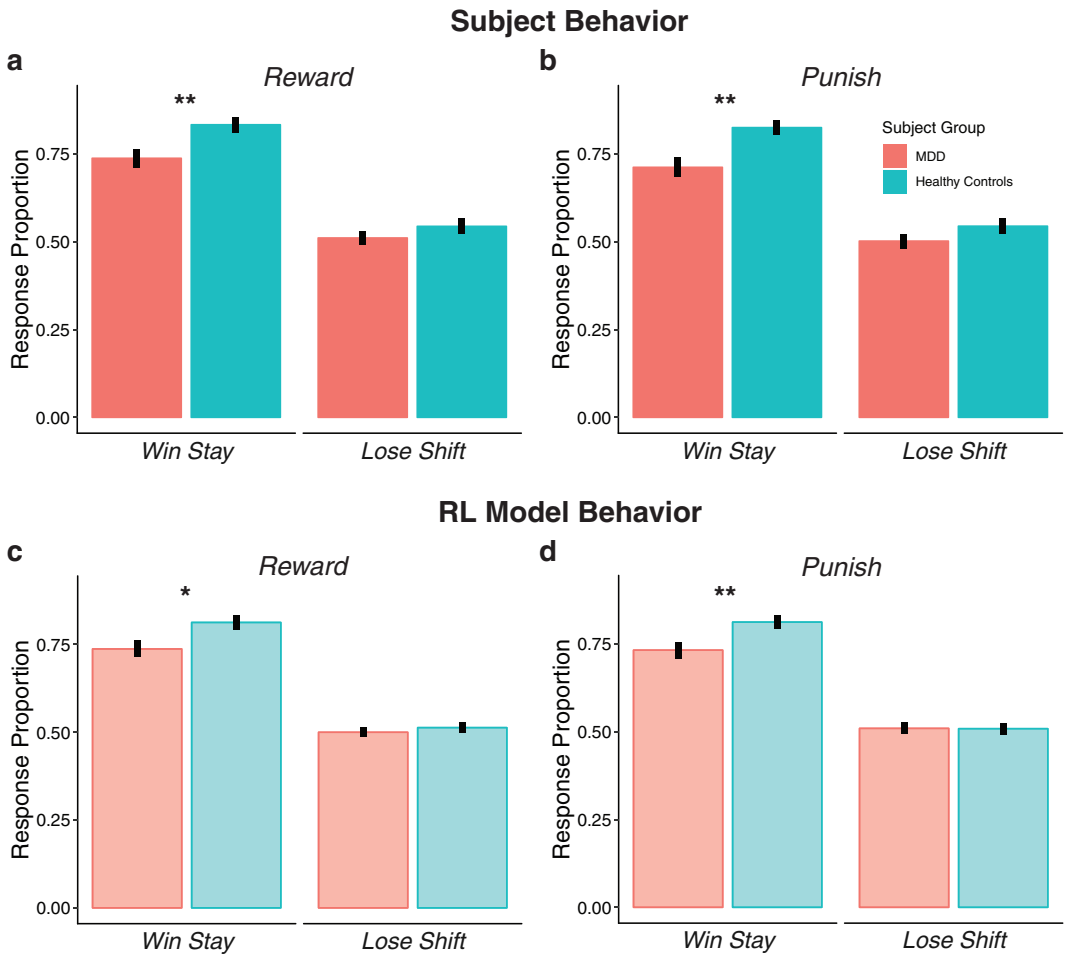

Figure 3. Individuals with MDD (red) made fewer 'win-stay' but similar 'lose-shift' choices compared to controls (blue) in both task conditions (a–b). Proportion 'win-stay' and 'lose-shift' responses by subjects in the reward (a) and punishment (b) conditions (c–d). Proportion 'win-stay' and 'lose-shift' responses from simulated behavior using subject parameter estimates (100 simulations per subject). Bars and errorbars correspond to means and one standard error of the mean respectively. MDD = major depressive disorder; RL = reinforcement learning. * $p < .05$. ** $p < .01$. See the online article for the color version of this figure.

There was also a clear influence of action learning that was most apparent in the 'win-stay' and 'lose-shift' behavior. In both the reward and punishment learning tasks, depressed individuals and controls were significantly more likely to make 'win-stay' responses when the same motor action repeated across trials ($ps <$ .05; Figure 4a–b). There was also a tendency for healthy controls to make more 'lose-shift' responses when the motor action for the previously selected fractal repeated across trials in the reward task, and a similar tendency in depressed individuals in the punishment task (all $ps < .05$).

Finally, we fit a logistic regression model (eq. 1) to estimate the influence of past outcomes on the current choice. Behavior consistent with reinforcement learning should show an exponential decrease in the influence of outcomes received from previous trials, as beliefs are changed incrementally with each new outcome. As is evident from Figure 5, both groups showed this exponential decay. However, controls showed a steeper decay of influence from previous outcomes, with choices depending on outcomes received up to two trials back, whereas depressed individuals were influenced by outcomes received up to three trials back. Such a steeper decay is consistent with higher learning-rates in healthy controls.

Depressed Individuals had Lower Learning-Rates and Lower Value Sensitivity Than Controls

We fit a reinforcement learning model to each subject's behavior and compared parameter values between groups to identify the learning components affected in depressed individuals. The main learning components in the model were the learning-rate and value sensitivity for the fractals (α_f and β_f), which capture how much subjects incorporate trial-by-trial feedback, and how sensitive they are to fractal value differences, respectively. Our model also included parameters that capture biases in subject performance, including action learning (β_a), perseverative biases (ρ_f and ρ_a), and choice biases (π_f and π_a). These parameters were fit to subject data

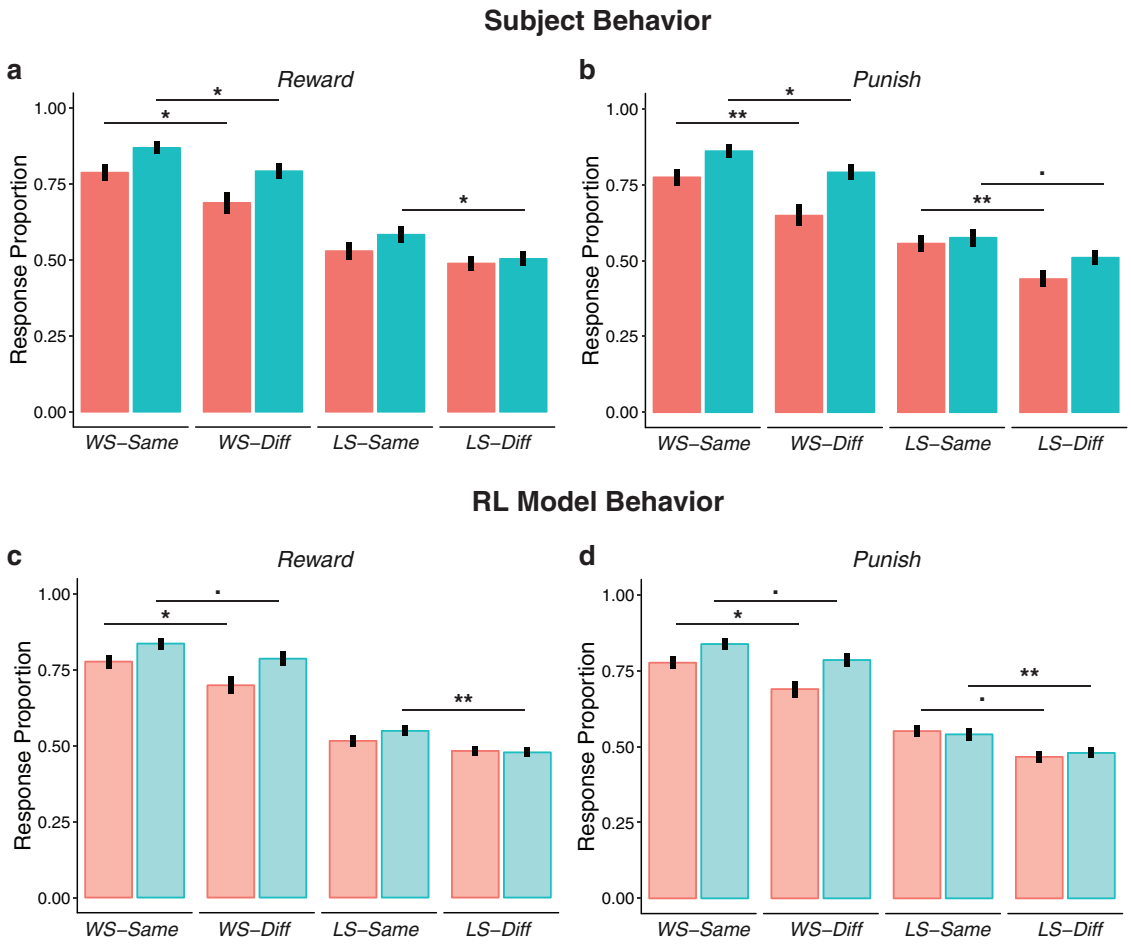

Figure 4. Both MDD (red) and healthy controls (blue) show action learning biases (a, b). Subject proportion "win-stay" (WS) and "lose-shift"'(LS) responses as a function of whether the fractals were on the same (Same) or different (Diff) side as the previous trial in both the reward (a) and punishment (b) conditions (c–d). Metrics from a, b for simulated behavior using subject parameter estimates (100 simulations per subject). Bars and errorbars correspond to means and one standard error of the mean respectively. RL = reinforcement learning. · $p < .1$. * $p < .05$. ** $p < .01$. See the online article for the color version of this figure.

using a hierarchical Bayesian model with group-level priors over each parameter (see methods). In evaluating model fits, each of the seven free parameters in the model captured a specific tendency in subject behavior (Supplementary Figure 1). Moreover, a model that included all terms better fit subject performance than reduced models that omitted action learning, perseveration, and/or bias terms (see Table 2).

These model fits captured all of the aspects of task performance and the performance differences between the two groups described above. Model simulations produced using subject parameters reproduced overall learning differences between the depressed and control groups, as well as differences in learning dynamics after reversals (Figure 2c–d). Model simulations also recapitulated the difference in proportion of win-stay choices, and the lack of difference in lose-shift choices, between depressed individuals and healthy controls (Figure 3c–d). In addition, model simulations further captured differences in the proportion of win-stay choices when the motor action repeated (WS-Same) compared to when it did not (WS-Diff; Figure 4).

These model fits revealed that depressed individuals had both lower learning-rates and slightly reduced value sensitivities. In both the reward and the punishment conditions, depressed subjects had lower fractal learning-rates (signed-rank test for differences in α_f - Reward: $p = .005$, Punishment: $p = .0007$). The two subject groups also differed in their fractal value sensitivity (β_f), with a trending difference in the reward condition ($p = .085$) and a larger difference in the punishment condition ($p = .04$). Beyond these major differences, controls also differed from depressed subjects in their fractal bias parameter (ρ_f) in the punishment condition ($p = .01$), which captured a slight but nonsignificant tendency to choose the second fractal more than the first fractal (mean proportion second fractal choices in punishment condition [*SE*]: Controls = .54 [.01], MDD = .51 [.02]; $p = .15$). The two groups did not differ on any other parameters in either condition (all ps $> .148$; Figure 6).

Within the MDD group, model parameters did not consistently differ between individuals undergoing different forms of treatment. Individuals treated with medication only tended to have

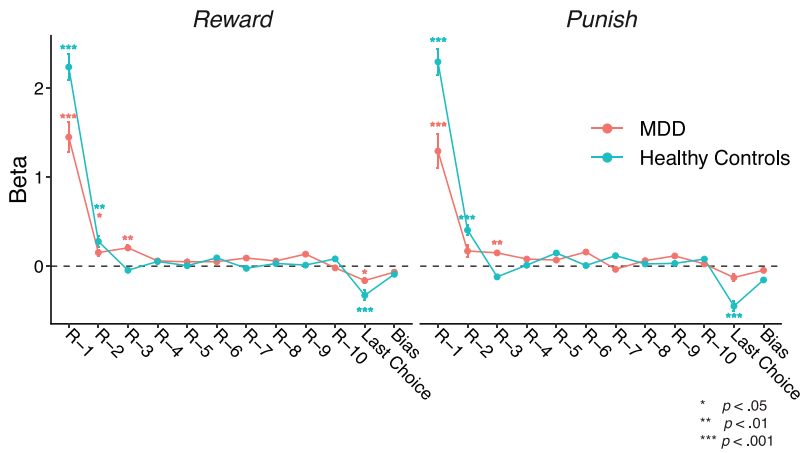

Figure 5. Both subject groups show learning dynamics consistent with reinforcement learning. Beta weights from a logistic regression (eq. 1) measuring the influence of positive outcomes 1–10 trials back (R-1 – R-10), the last fractal choice, and a bias toward either fractal in both the reward and punishment conditions. MDD = major depressive disorder. See the online article for the color version of this figure.

lower learning-rates than those not on medication in the reward condition (mean learning-rates [*SD*]: Medication only = 0.54 [0.17], No-medication = 0.69 [0.17]; signed-rank test, $p = .041$), but not in the punishment condition ($p = .419$). No other parameters in either condition differed between individuals being treated with medication only versus no medication, between individuals undergoing therapy only versus no therapy, or between individuals under any treatment (therapy and/or medication) versus no treatment (no medication or therapy; all signed-rank ps $> .05$).

Consistent with parameter differences between depressed individuals and healthy controls, fractal learning-rate (α_f) in both the reward and punishment conditions was negatively correlated with depression severity across the entire sample (Supplementary Table 2), as well as with other self-report dimensions that differed between the two groups, most notably anxiety levels and cognitive biases (e.g., avoidance, negative self-esteem). We additionally found correlations between clinical dimensions and action bias (π_a) in the reward condition and fractal bias (π_f) in the punishment condition. However, we did not see any consistent significant correlations with fractal sensitivity (β_f) or with any other model parameters in either condition.

Table 2
Model Comparison Between Full Model (First Row) and Reduced Models Using Approximate Leave-One-Out Cross Validation

Model parameters	Reward		Punish	
	LOOIC	ELPD (*SE*)	LOOIC	ELPD (*SE*)
α_f, β_f, β_a, ρ_f, ρ_a, π_f, π_a	11586	−5793 (160)	11686	−5843 (171)
α_f, β_f, β_a, ρ_f, ρ_a	11632	−5816 (160)	11760	−5880 (172)
α_f, β_f, β_a	11852	−5926 (159)	11958	−5979 (168)
α_f, β_f	12182	−6091 (166)	12222	−6111 (175)

Note. LOOIC = leave-one-out information criterion (lower values indicate better fit); ELPD = expected log predictive density (higher values indicate better model fit); *SE* = ELPD standard error.

Model Parameters Classify Depressed Versus Controls Better Than Simpler Metrics of Performance

Finally, we examined whether computational model fits improved the ability to correctly identify depressed individuals from controls, over and above more basic measures of task performance. To account for multicollinearity across parameters and behavioral metrics, we first used LASSO regression to identify the model parameters and behavioral metrics best able to classify depressed individuals from controls in each task condition (see methods). A LASSO regression with model parameters alone identified fractal learning-rate (α_f) and value sensitivity (β_f) as the most predictive parameters in both the reward and punishment conditions (i.e., regression weights >0), along with fractal bias (π_f) in the punishment condition. A LASSO regression with behavioral metrics alone identified proportion rich fractal choices, proportion win-stay responses, and action repeats (e.g., action perseveration) as the most predictive metrics in the reward condition, whereas only proportion rich fractal choices was identified in the punishment condition. When all model parameters and behavioral metrics were added in the same LASSO regression, fractal learning-rate, sensitivity for actions (β_a), proportion rich fractal choices, and fractal and action repetitions were most predictive in the reward condition, whereas fractal learning-rate, fractal bias, and proportion rich fractal choices were most predictive in the punishment condition.

We next input the model parameters and behavioral metrics identified by the LASSO regression into a logistic classifier to assess the ability of these variables to predict depressed status. In the reward condition, out-of-sample classification accuracy was highest for a logistic classifier trained using all of the identified model parameters and behavioral metrics in the reward condition, and this classifier provided better accuracy than one using any of the individual model parameters or behavioral metrics alone (AUC

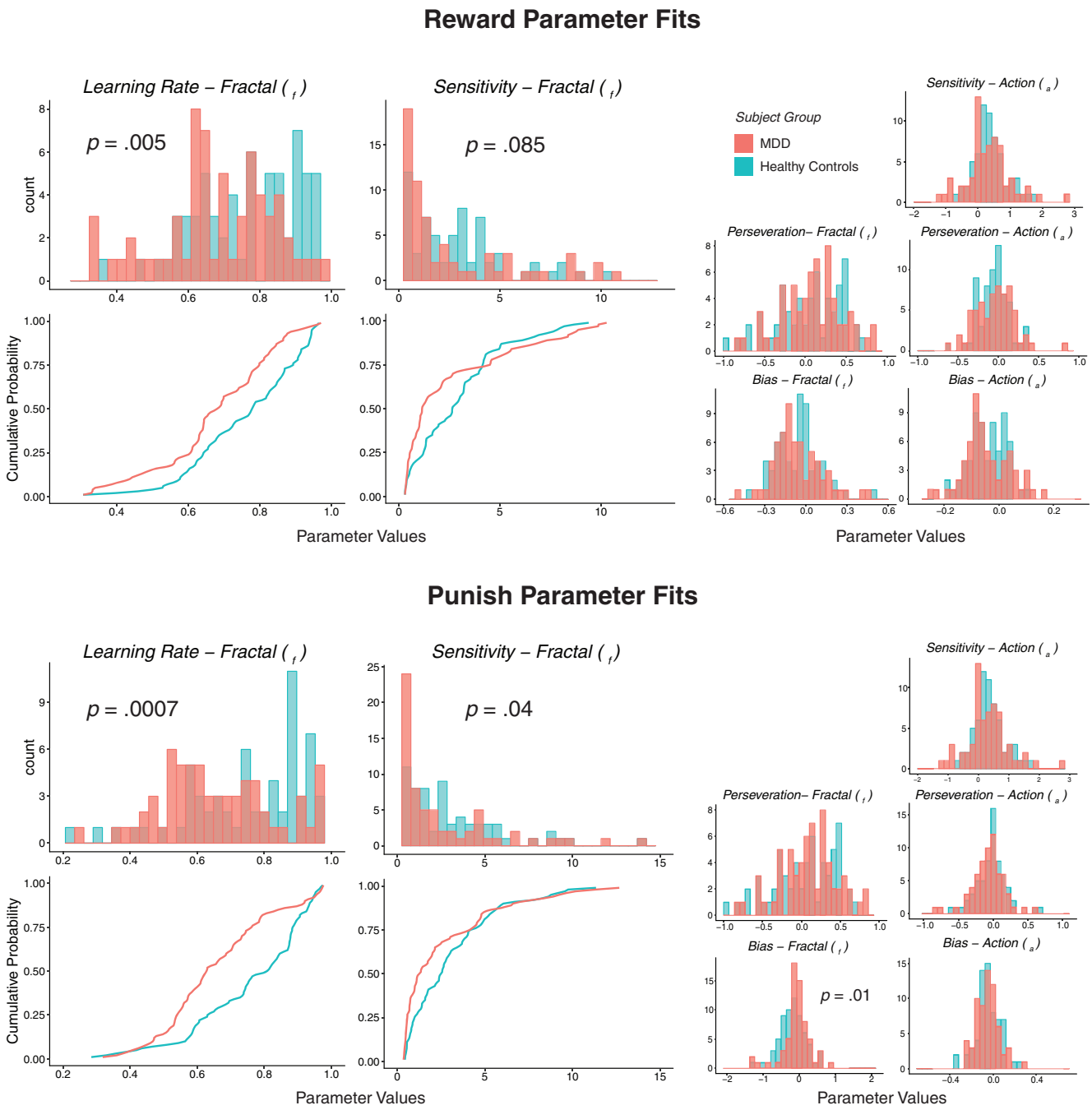

Figure 6. Individuals with major depressive disorder (MDD; red) have lower learning-rates and value sensitivity than healthy controls (blue). Histograms and respective cumulative probability plots of individual parameter values fits in the reward and punishment conditions. *p* values correspond to Wilcoxon signed-ranks test between subject groups. See the online article for the color version of this figure.

for full model = .69; Delong test comparing full model AUCs with other models: all *p*s < .03; Figure 7a). In the punishment condition, the full model provided better classification accuracy than any of the individual model parameters or behavioral metrics (*p*s < .05), except for learning-rate ($p = .121$; Figure 7b).

In both the reward and punishment conditions, learning-rates and proportion rich responses were the most consistent predictors of depression. We ran a final logistic classifier in both conditions that only included these two variables to assess whether they performed at the level of the full classifier models. Indeed, these reduced classifiers performed as well as the full model classifiers (AUC – Reward Condition: full model = .69, reduced model = .69; $p = .464$; Punishment Condition: full model = .70, reduced model = .71; $p = .730$; Figure 7a,7b).

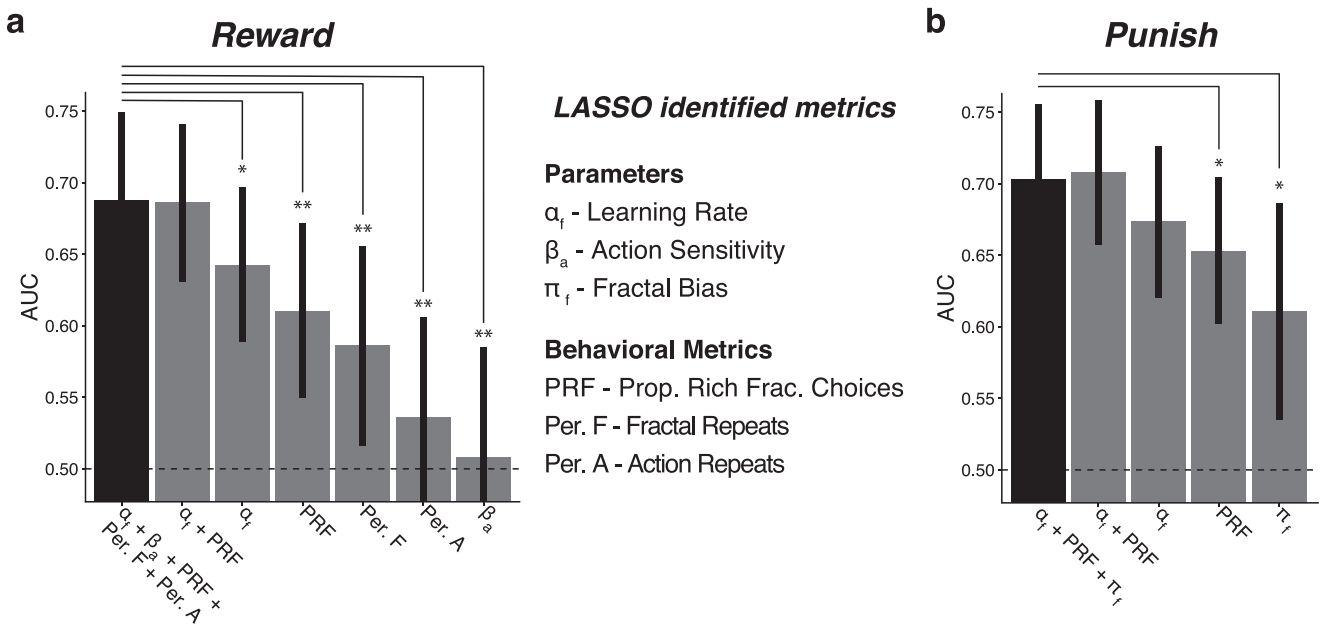

Figure 7. Learning-rate improves out-of-sample depression classification accuracy beyond simple performance metrics in both task conditions (a, b). Out-of-sample area under the receiver operating characteristic curve (AUC) for logistic regression models with model parameter and behavioral metrics that survived LASSO regressions for the reward (a) and punishment conditions (b). Bars and errorbars correspond to median and bootstrapped standard error estimates of AUC values (2000 bootstrapped samples). * $p < .05$. ** $p < .01$.

Discussion

This study sought a mechanistic understanding as to why depressed individuals learn less effectively from rewards and punishments than controls. Overall, we found that depressed individuals learn more slowly from feedback than healthy individuals and appear less sensitive to outcomes, as evidenced by lower win-stay response rates, shallower learning curves after reversals, shallower exponential decays from previous reward, and lower fit learning-rate and value sensitivity parameters. These differences held irrespective of whether the task involved learning from rewards or punishments, suggesting that group differences in both conditions arise from similar mechanisms. Furthermore, our modeling approaches help rule out the influence of other potential biases (i.e., perseverative, choice, or action biases) on the group differences we report.

Our findings are consistent with previous work showing reduced responsiveness to reward in depressed individuals (Eshel & Roiser, 2010; Henriques & Davidson, 2000; Henriques, Glowacki, & Davidson, 1994; Pizzagalli et al., 2008, 2005; Rupprechter, Stankevicius, Huys, Steele, & Seriès, 2018). Evidence for deficits in probabilistic reward learning has been somewhat inconsistent, though, with some previous studies failing to find overall depression-related differences (Chase et al., 2010; Dombrovski et al., 2010; Dombrovski et al., 2013; Dombrovski et al., 2015; Moutoussis et al., 2018). This discrepancy likely results from lower statistical power. Indeed, our study has twice the sample size of any previous study (Chase et al., 2010; Dombrovski et al., 2010; Dombrovski et al., 2013; Dombrovski et al., 2015;; Moutoussis et al., 2018; Rupprechter et al., 2018), and in several previous studies the depressed groups showed similar, albeit nonsignificant, trends toward poorer performance and lower learning-rates (Chase et al., 2010; Dombrovski et al., 2010). As such, our results support the hypothesis that depression leads to reductions in reward-related learning.

However, our results do not support the hypothesis of a global increase in sensitivity to punishment, with depressed individuals showing similar performance deficits when learning to avoid punishment. Moreover, depressed individuals did not differ from controls in their level of 'lose-shift' responses. Evidence for increased 'lose-shift' responses in depressed individuals has been inconsistent, with some studies showing heightened responses in depressed subjects (Dombrovski et al., 2013; Murphy et al., 2003) and others not (Chase et al., 2010; Chen et al., 2015; Dombrovski et al., 2010; Gradin et al., 2011). Previous studies have shown that rewarding feedback (e.g., rewarded trials in our reward condition) is treated similarly to lack of punishment in aversion learning tasks (e.g., unpunished trials in the punishment condition), at both the behavior and neural levels (Erdeniz & Done, 2019; Palminteri et al., 2012). As such, our detailed analysis supports the idea that depression results in a global reduction in sensitivity to positive outcomes, regardless of the context in which the positive outcome is delivered (e.g., even if lack of punishment serves as the positive outcome).

Although our results do not support global asymmetries in the ways depressed individuals learn from rewards and punishments, responses to rewards and punishments may still differ in specific contexts. Evidence for hypersensitivity to negative feedback has been found primarily in more cognitive tasks, and/or under conditions where negative feedback is social in nature (e.g., from the experimenters administering the task; Beats et al., 1996; Elliott et

al., 1996; Eshel & Roiser, 2010). While our results do not support a general hypersensitivity to negative feedback in MDD, such a tendency might instead occur in more specific cognitive or social contexts.

Our computational analysis also provides insights into the feedback processing mechanisms affected by depression. Learning-rates were consistently lower in individuals with MDD in both task conditions. Recent work has proposed that, rather than decreased learning-rates, reward processing deficits in depression arise from lower value sensitivity, a factor not often examined in many RL analyses of MDD behavior (Huys et al., 2013; Huys et al., 2015). Consistent with this claim, we also observe that value sensitivity is reduced in MDD, particularly in punishment learning. However, our out-of-sample prediction analysis suggests that these sensitivity differences contribute less than learning-rates do to MDD deficits in reversal-learning.

We note two important caveats regarding these findings on learning-rate and value sensitivity. First, in our task, changes in value sensitivity have the same effect in an RL model as changes in the exploration-exploitation tradeoff. Therefore, the reduced value sensitivity we see in depressed individuals could be interpreted as either a greater tendency toward exploratory choices or a reduced subjective desirability of positive versus negative outcomes. Second, task conditions in which depressed individuals show lower reward sensitivity rather than lower learning-rate in previous studies (Huys et al., 2013) involve a qualitatively different environment, where subjects make perceptual judgments that are implicitly influenced by rewards observed on previous trials (Pizzagalli et al., 2008). It is possible that in explicit learning environments, such as reversal-learning tasks, the influence on learning-rate becomes more apparent. Indeed, consistent with our findings, a more recent study using a task similar to our reward condition also reports lower learning-rate and value sensitivity in depressed individuals, suggesting deficits in both computations (Rupprechter et al., 2018).

Recent views have proposed that learning deficits in MDD may be explained by deficits in learning the underlying structure of probabilistic tasks (e.g., identifying the underlying fractal reward probabilities, which is proposed to be a function of the "model-based" valuation system; Huys et al., 2015). This is certainly a possibility, and the fact that we find similar learning deficits in reward and punishment contexts lends support to this notion. However, since state estimations in our task are directly related to outcome history, dissociating the specific contributions of outcome processing versus state estimations in our task is challenging. Future research could address differences in outcome processing versus state estimations using tasks better designed to tease these two apart (e.g., two-step tasks; Daw et al., 2011).

Although not measured directly in the current study, our results also suggest neurobiological implications. Reinforcement learning has been linked to dopaminergic signals in the striatum (Daw et al., 2011; O'Doherty et al., 2004; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006). A number of studies have found that striatal responses to reward, including those received during reversal-learning tasks, were attenuated in depressed individuals (Gradin et al., 2011; Kumar et al., 2008; Pizzagalli et al., 2009; Remijnse et al., 2009; Robinson, Cools, Carlisi, Sahakian, & Drevets, 2012). Another study found that reduced ventral striatal responsiveness to unexpected rewards predicted the severity of depression across both unipolar and bipolar depressed groups (Satterthwaite et al., 2015). These results are consistent with MDD blunting responsiveness to positive outcomes, which could subsequently influence how feedback is used for learning. A key priority for future research should be to more closely link these neural and behavioral effects of depression, and to explore these effects across a wider range of disorders where blunted value responsiveness is a key component, including across mood (unipolar and bipolar depression) and psychotic disorders (Gold, Waltz, Prentice, Morris, & Heerey, 2008; Waltz, Frank, Robinson, & Gold, 2007).

The current results need to be appraised in light of certain limitations. Although we did screen out individuals with bipolar, substance abuse and/or psychotic symptoms, we did not assess other potential comorbidities. Beyond the screen for substance use disorder, we did not collect detailed information on alcohol use or smoking, which may be associated with learning performance.

Additionally, previous research has shown that anxiety disorders also affect learning (Browning, Behrens, Jocham, O'Reilly, & Bishop, 2015; Huang, Thompson, & Paulus, 2017), and we find that learning-rates are negatively correlated with anxiety severity in our sample. Depressive and anxiety disorders are highly comorbid, and separating their distinct contributions to learning was beyond the scope of the current study. However, differences between our results and previous studies suggest some possible distinctions between depression and anxiety. Individuals with higher trait anxiety show a reduced ability to adapt to contextual task changes (i.e., to stable vs. volatile environments), but only in tasks where subjects learn from punishment (Browning et al., 2015). Similarly, high anxiety has been associated with increases in "lose-shift" responses, again suggesting anxiety may specifically impact learning from negative feedback (Huang et al., 2017). In contrast, here we found decreases in "win-stay" but not "lose-shift" responses, suggesting these behavioral markers may differentiate between learning impairments resulting from depression or anxiety. A priority for future research should be to further distinguish the separate impacts of anxiety and depression on learning.

Finally, the current cross-sectional study does not address whether impaired reward and punishment learning is a risk factor for developing depression or a consequence of the disorder. Our results suggest that type of treatment (therapy and/or medication) does not significantly impact reversal-learning behavior. Future studies could further address these questions through longitudinal or treatment studies. Whether individuals who exhibit poorer learning are more likely to develop depression, and/or whether these deficits increase with the onset of depressive symptoms and improve in remission, are both critical questions for understanding the association between depression and reward/punishment learning.

References

Beats, B. C., Sahakian, B. J., & Levy, R. (1996). Cognitive performance in tests sensitive to frontal lobe dysfunction in the elderly depressed. *Psychological Medicine, 26,* 591–603. http://dx.doi.org/10.1017/S0033291700035662

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56,* 893–897. http://dx.doi.org/10.1037/0022-006X.56.6.893

Beck, A., Steer, R., & Brown, G. (1996). *Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Betancourt, M., & Girolami, M. (2015). Hamiltonian Monte Carlo for hierarchical models. In S. K. Upadhyay, U. Singh, D. K. Dey, & A. Loganathan (Eds.), *Current trends in Bayesian methodology with applications* (pp. 79–101). Boca Raton, FL: CRC Press.

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience, 18,* 590–596. http://dx.doi.org/10.1038/nn.3961

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*. Advance online publication. http://dx.doi.org/10.18637/jss.v076.i01

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67,* 319–333. http://dx.doi.org/10.1037/0022-3514.67.2.319

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,* 433–440. http://dx.doi.org/10.1017/S0033291709990468

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: A review of computational research. *Neuroscience and Biobehavioral Reviews, 55,* 247–267. http://dx.doi.org/10.1016/j.neubiorev.2015.05.005

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69,* 1204–1215. http://dx.doi.org/10.1016/j.neuron.2011.02.027

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44,* 837–845. http://dx.doi.org/10.2307/2531595

Dombrovski, A. Y., Clark, L., Siegle, G. J., Butters, M. A., Ichikawa, N., Sahakian, B. J., & Szanto, K. (2010). Reward/Punishment reversal learning in older suicide attempters. *The American Journal of Psychiatry, 167,* 699–707. http://dx.doi.org/10.1176/appi.ajp.2009.09030407

Dombrovski, A. Y., Szanto, K., Clark, L., Aizenstein, H. J., Chase, H. W., Reynolds, C. F., III, & Siegle, G. J. (2015). Corticostriatothalamic reward prediction error signals and executive control in late-life depression. *Psychological Medicine, 45,* 1413–1424. http://dx.doi.org/10.1017/S0033291714002517

Dombrovski, A. Y., Szanto, K., Clark, L., Reynolds, C. F., & Siegle, G. J. (2013). Reward signals, attempted suicide, and impulsivity in late-life depression. *Journal of the American Medical Association Psychiatry,* Advance online publication. http://dx.doi.org/10.1001/jamapsychiatry.2013.75

Elliott, R., Sahakian, B. J., McKay, A. P., Herrod, J. J., Robbins, T. W., & Paykel, E. S. (1996). Neuropsychological impairments in unipolar depression: The influence of perceived failure on subsequent performance. *Psychological Medicine, 26,* 975–989. http://dx.doi.org/10.1017/S0033291700035303

Erdeniz, B., & Done, J. (2019). Common and Distinct Functional Brain Networks for Intuitive and Deliberate Decision Making. *Brain Sciences, 9,* 174. http://dx.doi.org/10.3390/brainsci9070174

Eshel, N., & Roiser, J. P. (2010). Reward and punishment processing in depression. *Biological Psychiatry, 68,* 118–124. http://dx.doi.org/10.1016/j.biopsych.2010.01.027

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured clinical interview for DSM–IV–TR axis I disorders, research version, patient edition. (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute.

Gelman, A., & Rubin, D. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science, 7,* 457–511. http://dx.doi.org/10.1214/ss/1177011136

Gold, J. M., Waltz, J. A., Prentice, K. J., Morris, S. E., & Heerey, E. A. (2008). Reward processing in schizophrenia: A deficit in the representation of value. *Schizophrenia Bulletin, 34,* 835–847. http://dx.doi.org/10.1093/schbul/sbn068

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., . . . Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain: A Journal of Neurology, 134*(Pt. 6), 1751–1764. http://dx.doi.org/10.1093/brain/awr059

Henriques, J. B., & Davidson, R. J. (2000). Decreased responsiveness to reward in depression. *Cognition and Emotion, 14,* 711–724. http://dx.doi.org/10.1080/026999300501176 84

Henriques, J. B., Glowacki, J. M., & Davidson, R. J. (1994). Reward fails to alter response bias in depression. *Journal of Abnormal Psychology, 103,* 460–466. http://dx.doi.org/10.1037/0021-843X.103.3.460

Huang, H., Thompson, W., & Paulus, M. P. (2017). Computational dysfunctions in anxiety: Failure to differentiate signal from noise. *Biological Psychiatry, 82,* 440–446. http://dx.doi.org/10.1016/j.biopsych.2017.07.007

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Computational Biology, 7*(4), e1002028. http://dx.doi.org/10.1371/journal.pcbi.1002028

Huys, Q. J. M., Daw, N. D., & Dayan, P. (2015). Depression: A decision-theoretic analysis. *Annual Review of Neuroscience, 38,* 1–23. http://dx.doi.org/10.1146/annurev-neuro-071714-033928

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience, 19,* 404–413. http://dx.doi.org/10.1038/nn.4238

Huys, Q. J. M., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders, 3,* 12. http://dx.doi.org/10.1186/2045-5380-3-12

Keung, W., Hagen, T. A., & Wilson, R. C. (2019). Regulation of evidence accumulation by pupil-linked arousal processes. *Nature Human Behaviour, 3,* 636–645. http://dx.doi.org/10.1038/s41562-019-0551-4

Kuhn, M. (2008). Introduction. *Journal of Statistical Software, 28,* 1–26.

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain: A Journal of Neurology, 131*(Pt. 8), 2084–2093. http://dx.doi.org/10.1093/brain/awn136

Lau, B., & Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. *Neuron, 58,* 451–463. http://dx.doi.org/10.1016/j.neuron.2008.02.021

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33,* 335–343. http://dx.doi.org/10.1016/0005-7967(94)00075-U

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

Moutoussis, M., Rutledge, R. B., Prabhu, G., Hrynkiewicz, L., Lam, J., Ousdal, O. T., . . . Dolan, R. J. (2018). Neural activity and fundamental learning, motivated by monetary loss and reward, are intact in mild to moderate major depressive disorder. *PLoS ONE, 13*(8), e0201451. http://dx.doi.org/10.1371/journal.pone.0201451

Mukherjee, D., Lee, S., Kazinka, R. D., Satterthwaite, T., & Kable, J. W. (2020). Multiple facets of value-based Decision Making in Major Depressive Disorder. *Scientific Reports, 10,* 3415. http://dx.doi.org/10.1038/s41598-020-60230-z

Murphy, F. C., Michael, A., Robbins, T. W., & Sahakian, B. J. (2003). Neuropsychological impairment in patients with major depressive disorder: The effects of feedback on task performance. *Psychological Medicine, 33,* 455–467. http://dx.doi.org/10.1017/S0033291702007018

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304,* 452–454. http://dx.doi.org/10.1126/science.1094285

Ottenbreit, N. D., & Dobson, K. S. (2004). Avoidance and depression: The construction of the Cognitive-Behavioral Avoidance Scale. *Behaviour research and therapy, 42,* 293–313. http://dx.doi.org/10.1016/S0005-7967(03)00140-2

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., . . . Pessiglione, M. (2012). Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron, 76,* 998–1009. http://dx.doi.org/10.1016/j.neuron.2012.10.017

Pechtel, P., Dutra, S. J., Goetz, E. L., & Pizzagalli, D. A. (2013). Blunted reward responsiveness in remitted depression. *Journal of Psychiatric Research, 47,* 1864–1869. http://dx.doi.org/10.1016/j.jpsychires.2013.08.011

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442,* 1042–1045. http://dx.doi.org/10.1038/nature05051

Pizzagalli, D. A., Holmes, A. J., Dillon, D. G., Goetz, E. L., Birk, J. L., Bogdan, R., . . . Fava, M. (2009). Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *The American Journal of Psychiatry, 166,* 702–710. http://dx.doi.org/10.1176/appi.ajp.2008.08081201

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,* 76–87. http://dx.doi.org/10.1016/j.jpsychires.2008.03.001

Pizzagalli, D. A., Jahn, A. L., & O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry, 57,* 319–327. http://dx.doi.org/10.1016/j.biopsych.2004.11.026

Razani, J., Murcia, G., Tabares, J., & Wong, J. (2007). The effects of culture on WASI test performance in ethnically diverse individuals. *The Clinical Neuropsychologist, 21,* 776–788. http://dx.doi.org/10.1080/13854040701437481

Remijnse, P. L., Nielen, M. M. A., van Balkom, A. J. L. M., Hendriks, G. J., Hoogendijk, W. J., Uylings, H. B. M., & Veltman, D. J. (2009). Differential frontal-striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive-compulsive disorder. *Psychological Medicine, 39,* 1503–1518. http://dx.doi.org/10.1017/S0033291708005072

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 8,* 12. http://dx.doi.org/10.1186/1471-2105-12-77

Robinson, O. J., Cools, R., Carlisi, C. O., Sahakian, B. J., & Drevets, W. C. (2012). Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. *The American Journal of Psychiatry, 169,* 152–159. http://dx.doi.org/10.1176/appi.ajp.2011.11010137

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rupprechter, S., Stankevicius, A., Huys, Q. J. M., Steele, J. D., & Seriès, P. (2018). Major depression impairs the use of reward values for decision-making. *Scientific Reports, 8,* 13798. http://dx.doi.org/10.1038/s41598-018-31730-w

Satterthwaite, T. D., Kable, J. W., Vandekar, L., Katchmar, N., Bassett, D. S., Baldassano, C. F., . . . Wolf, D. H. (2015). Common and dissociable dysfunction of the reward system in bipolar and unipolar depression. *Neuropsychopharmacology, 40,* 2258–2268. http://dx.doi.org/10.1038/npp.2015.75

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *The British Journal of Psychiatry, 167,* 99–103. http://dx.doi.org/10.1192/bjp.167.1.99

Steffens, D. C., Wagner, H. R., Levy, R. M., Horn, K. A., & Krishnan, K. R. R. (2001). Performance feedback deficit in geriatric depression. *Biological Psychiatry, 50,* 358–363. http://dx.doi.org/10.1016/S0006-3223(01)01165-9

Sutton, R., & Barto, A. (1998). *Introduction to reinforcement learning*. Cambridge, MA: MIT Press. http://dx.doi.org/10.1109/TNN.1998.712192

Stan Development Team. (2018). RStan: The R interface for Stan. Retrieved from https://mc-stan.org/rstan/articles/rstan.html

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27,* 1413–1432. http://dx.doi.org/10.1007/s11222-016-9696-4

Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological Psychiatry, 62,* 756–764. http://dx.doi.org/10.1016/j.biopsych.2006.09.042

Wang, X.-J., & Krystal, J. H. (2014). Computational psychiatry. *Neuron, 84,* 638–654. http://dx.doi.org/10.1016/j.neuron.2014.10.018

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence (WASI-II; 2nd ed.)*. San Antonio, TX: NCS Pearson.

Received February 12, 2020
Revision received July 31, 2020
Accepted August 3, 2020 ■