Aberrant Aversive Learning Signals in the Habenula in Remitted Patients With Recurrent Depression

de Klerk-Sluis, Jessica M.; Geugies, Hanneke; Mocking, Roel J.T.; Figueroa, Caroline A.; Groot, Paul F.C.; Marsman, Jan Bernard C.; van Eijndhoven, Philip F.P.; Geurts, Dirk E.M.; Ruhé, Henricus G.

DOI
10.1016/j.bpsc.2025.04.006

Publication date
2025

Document Version
Final published version

Published in
Biological Psychiatry: Cognitive Neuroscience and Neuroimaging

Biological Psychiatry: CNNI

Archival Report

Aberrant Aversive Learning Signals in the Habenula in Remitted Patients With Recurrent Depression

Jessica M. de Klerk-Sluis, Hanneke Geugies, Roel J.T. Mocking, Caroline A. Figueroa, Paul F.C. Groot, Jan-Bernard C. Marsman, Philip F.P. van Eijndhoven, Dirk E.M. Geurts, and Henricus G. Ruhé

ABSTRACT

BACKGROUND: Hypersensitivity to punishment is one of the core features of major depressive disorder (MDD). Hypersensitivity to punishment has been proposed to originate from aberrant aversive learning. One of the key areas in aversive learning is the habenula. Although evidence for dysfunctional aversive learning in patients with depression is well established, whether this dysfunction and its neural correlates persist during symptomatic remission of depression remains largely unexplored.

METHODS: Functional magnetic resonance imaging data from 36 medication-free remitted patients with recurrent MDD and 27 healthy control participants participating in a Pavlovian classical conditioning task were assessed within a computational modeling framework to evaluate temporal difference–related activation of the habenula during aversive learning. Furthermore, generalized psychophysiological interaction analyses were performed to assess functional connectivity of the temporal difference signal with the habenula as an a priori region of interest.

RESULTS: Relative to healthy control participants, patients showed significantly increased temporal difference–related aversive learning activation in the bilateral habenula. This activation was correlated with residual symptoms in the remitted MDD group. Furthermore, patients exhibited decreased functional connectivity between the habenula and the ventral tegmental area compared with control participants.

CONCLUSIONS: The increased habenula activity during aversive learning, particularly during the expectation of punishment, together with decreased functional habenula–ventral tegmental area connectivity in remitted patients with MDD, reflect hypersensitivity to and/or inability to regulate the impact of aversive environmental cues and punishment.

https://doi.org/10.1016/j.bpsc.2025.04.006

Major depressive disorder (MDD) poses a significant societal and financial burden, with recurrence rates reaching up to 80% within 5 years (1) and annual costs exceeding $326 billion in the United States (2). This necessitates prediction and prevention of recurrence, which requires identification and understanding of underlying pathophysiological mechanisms. MDD is characterized by hypersensitivity to stress, particularly punishment (3,4). This hypersensitivity often persists as a residual symptom after remission (5). It has been proposed that hypersensitivity to punishment originates from aberrant aversive learning (6).

Behavioral literature has demonstrated that the difference between the expectancy of an aversive stimulus and the actual outcome causes learning about the event. This difference can be captured by a concept called the prediction error (PE). Aversive PEs occur when an aversive stimulus is unexpectedly absent or less severe than anticipated or when an aversive stimulus is unexpectedly present or more severe than expected (7). In Pavlovian learning, PEs can be captured by a temporal difference (TD) learning algorithm (8). A meta-analysis of healthy individuals found aversive PE activity to be most prominent in the habenula and insula (7). In patients with MDD, habenula activation appears to be altered, showing either blunted responses during instrumental learning tasks (9,10) or even a negative relationship between the strength of aversive associations during a Pavlovian learning task and habenula blood oxygen level–dependent (BOLD) signal (11).

Given its central role in processing negative stimuli, the habenula is a key region of interest (ROI) in the study of MDD (12,13). A recent systematic review further supports the role of the habenula in depression, with 24 clinical studies having reported abnormalities in habenula connectivity, volume, and molecular markers (14). It has been suggested that in MDD, decreased serotonergic transmission elevates the activity of the habenula, which in turn inhibits the dopaminergic ventral tegmental area (VTA), thereby mediating depressive symptoms (15). Robust associations have been found between aberrant activity of the lateral habenula and depressive symptomatology

(16,17). Suppression of habenula hyperactivity has been shown to reduce depressive symptoms in rats (18). Furthermore, targeting habenula hyperactivity in rats by blocking habenula firing with ketamine elevated mood quickly and caused rapid relief of depressive symptoms (19). Moreover, deep brain stimulation in the lateral habenula caused remission of symptoms in a patient with treatment-resistant depression (20).

Because hypersensitivity to punishment and aversive learning abnormalities have been proposed as vulnerability markers for relapse, understanding habenula function in remitted MDD is crucial. Although evidence for a dysfunction in aversive learning in patients with MDD is well established (21), whether habenula dysfunction persists during remission remains largely unexplored. Furthermore, it is unknown whether alterations in functional connectivity between areas involved in aversive learning exist during remission. To address this gap in the literature, we evaluated, in remitted medication-free individuals with recurrent MDD, 1) TD-related activation of the habenula during a classical aversive condition functional magnetic resonance imaging (fMRI) task and 2) functional connectivity with the habenula as an a priori ROI. Based on work with currently depressed individuals, we hypothesized that there would be aberrant habenula activation and connectivity in the remitted MDD group compared with healthy control participants (HCs).

METHODS AND MATERIALS

Participants

The current study was part of a larger neuroimaging study investigating vulnerability for recurrence in MDD (between 2012 and 2016) (22). Participants ages 35 to 65 years with a known recurrent depressive disorder, currently in stable remission without medication, were recruited by advertisements and through contact with previous clinical treatment and/or previous studies. HCs were recruited via advertisements. All participants gave written informed consent. Permission for the study was obtained from the local ethics committee (22). Depression severity was assessed by an observer-rated 17-item Hamilton Depression Rating Scale ($HDRS_{17}$) (23), which was measured twice, once at intake and then again on the day of the MRI scan. The mean interval between assessments was 39.2 days and ranged from 1 to 129 days. Inclusion criteria for patients with MDD were 1) the presence of a recurrent depression defined as $\geq$2 MDD episodes according to a structured interview for DSM-IV (Structured Clinical Interview for DSM-IV [SCID]) and 2) stable remission defined as both an Inventory for Depressive Symptomatology—Self-Report (IDS-SR) score $\leq$14 and an $HDRS_{17}$ score $\leq$7 for at least 10 subsequent weeks. HCs were matched on sex, age, and years of education. Exclusion criteria for both groups were 1) psychotic or bipolar disorder, 2) current alcohol or drug dependence, 3) primary anxiety disorder, 4) MRI participation contraindications, 5) electroconvulsive therapy $<$2 months before scanning, and 6) a history of neurological disease. Furthermore, HCs were excluded if they had a psychiatric disorder (SCID) or first-degree relatives with a psychiatric disorder or if they scored $>$14 on the IDS-SR. All participants were free from psychopharmacological medication for $>$4 weeks. Of the 62 patients with remitted recurrent MDD (rrMDD) and 41 HCs initially scanned, 3 patients and 2 HCs were excluded due to structural abnormalities. Additionally, 5 patients and 4 HCs were excluded due to missing task data, and 18 patients and 8 HCs were excluded due to corrupt or missing physiological data. This resulted in a final sample of 36 patients and 27 HCs.

Task

A Pavlovian classical conditioning paradigm was used to assess aversive learning, utilizing a modified version of the task previously used by Kumar *et al.* (24). Participants were asked to abstain from liquids for 6 hours prior to scanning to ensure thirstiness. The task started with 1 block of 30 trials on which one of 2 pictures were presented (the to-be-conditioned stimuli [CS]) but without fluid delivery, as a neutral condition. Thereafter, the task contained 3 blocks of 30 trials, with each trial lasting for 8 seconds (Figure 1A). To create an aversive condition, 0.2 mL of bitter water (4 mol/L magnesium sulfate solution) was delivered as an unconditioned stimulus (US) at different probabilities. Two seconds after the start of a trial, one of 2 pictures was presented on the screen (CS), followed approximately 2 seconds later by the US. The probabilities of US delivery varied across blocks as follows: block 1: picture 1 = 80%, picture 2 = 20%; block 2: picture 1 = 20%, picture 2 = 80%; block 3: picture 1 = 80%, picture 2 = 20%. Before and after the task, participants received 0.2 mL of bitter water, after which they were asked to use a visual analog scale to rate how much they enjoyed/disliked the taste of the fluid (liking) and how much money they were willing to receive/pay to get more fluid (wanting). The bitter water was delivered through a polyethylene tube connected to a Braun-Infusomat pump interfaced with the stimulus computer. Stimuli were presented using E-prime 2 (Psychology Software Tools). Participants were instructed to try to find a pattern in which pictures predicted fluid delivery. It was explained that this association could change over time. TD signals were calculated based on the changing probabilities of fluid delivery (24).

TD Learning Model

For the TD analysis, we used a model-based fMRI approach incorporating a reinforcement learning algorithm (24,25). Each trial (i) was modeled consisting of 8 time points (t), where CS presentation was modeled at time point 3, and US delivery was modeled at time point 6. Based on previous literature, parameter settings were set as equal for all participants (11,21,26–28). For calculation of our TD error signal, we followed the procedure described by Geugies *et al.* (29). The predicted value (V) at any time point t was defined as

$$\hat{V}(t) = \sum_i w_i x_i(t) \quad (1)$$

where $x_i(t)$ is a vector with a 1 or a 0 (for all time points) representing the presence or absence of a CS at time point t. w_i represents a weight that was updated on a trial-by-trial basis in order to capture learning by

$$\Delta w_i = \alpha \sum_t x_i(t)\delta(t) \quad (2)$$

where α represents the learning rate. We selected plausible learning rates from the literature (0.4 and 0.5) and explored which learning rate fit our data best (11,26,30). For both

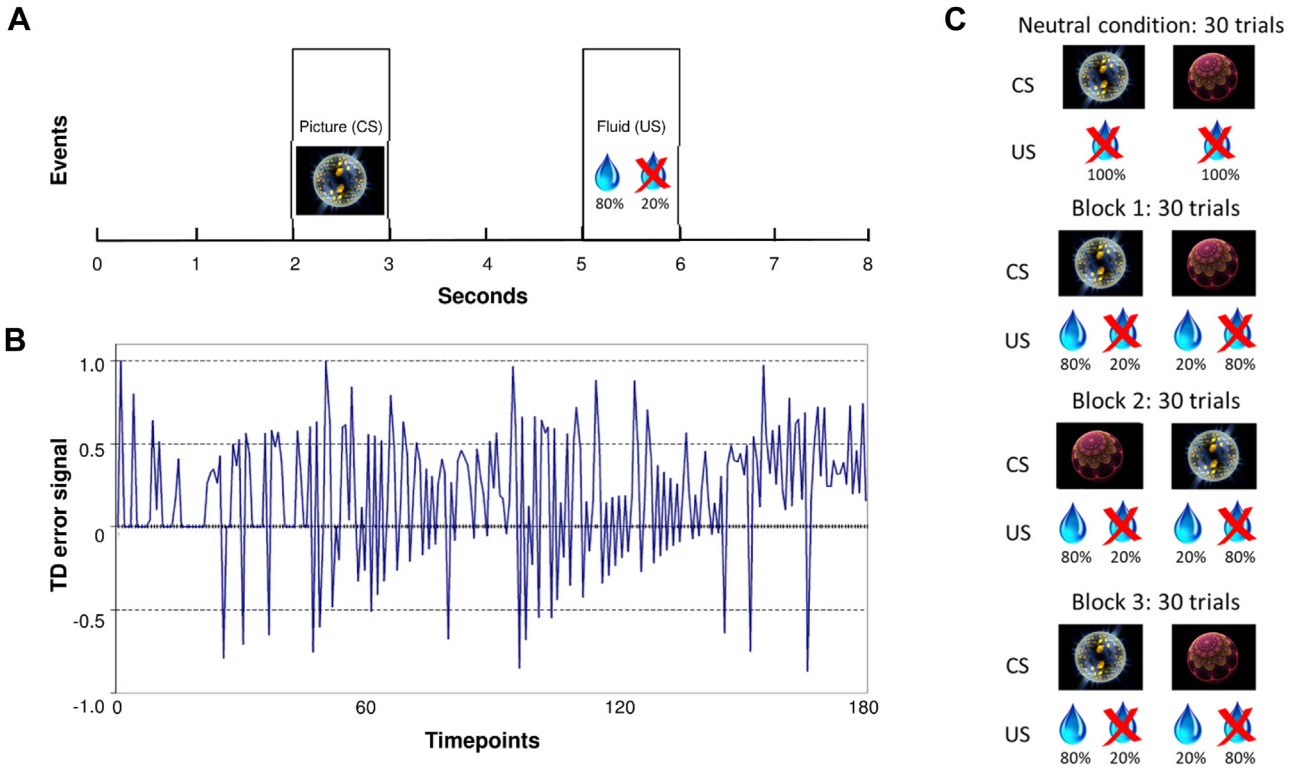

Figure 1. Pavlovian reinforcement task paradigm. **(A)** Timing of the conditioned stimulus (CS) and unconditioned stimulus (US) within 1 trial. **(B)** Example of a temporal difference (TD) error signal of one participant across 90 trials, showing 180 time points corresponding to the CS and US. **(C)** Visualization of the CS and US for the neutral condition and each of the 3 blocks.

learning rates, we calculated signal-to-noise values within our a priori habenula ROI. We also determined estimation efficiency values of statistical parametric mapping designs (31). Both methods supported a learning rate of $\alpha = .5$ (see Figure 2).

The TD error signal was defined as

$$\delta(t) = r(t) + \gamma \widehat{V}(t+1) - \widehat{V}(t) \quad (3)$$

where $r(t)$ is a vector with a 1 or a 0 (for all time points) for the presence or absence of bitter water, respectively and γ corresponds to a factor chosen in advance that determined the importance of later reinforcements compared with previous ones. Following previous studies, $\gamma = 1.0$ was used (24,32). Figure 1B shows an example of the TD error signal for 1 participant.

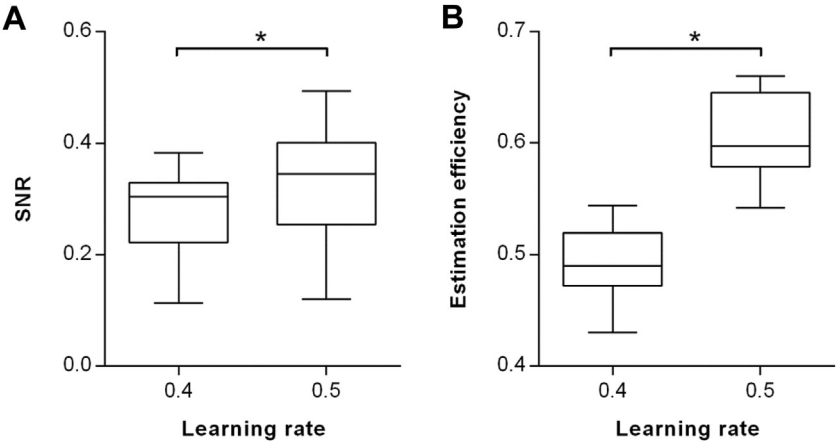

Figure 2. Model efficacy for different learning rates. **(A)** Signal-to-noise ratio (SNR) based on 1-group (all participants) contrast map. **(B)** Estimation efficiency of statistical parametric mapping designs across all participants. $^*p < .05$.

MRI Acquisition

MRI data were acquired with a 3T Philips Achieva XT scanner equipped with a 32-channel SENSE head coil. Functional images were acquired using a T2*-weighted gradient echo-planar imaging sequence. Imaging parameters were as follows: 25 slices (acquired in ascending order oriented with 30° tilt from the anterior commissure–posterior commissure transverse plane); TR = 1500 ms; TE = 28 ms; FOV = 240 × 240 mm and matrix = 80 × 80; voxel size = 3 mm; 1125 volumes. For anatomical reference, a high-resolution whole-brain T1-weighted image was acquired with the following parameters: 220 slices; TR = 8.3 ms; TE = 3.8 ms; FOV = 240 × 188 mm and matrix = 240 × 240; voxel size = 1 mm. During the scan, respiratory and cardiac signals were collected and were used in the analysis to correct for physiological noise.

MRI Data Processing and Analysis

Images were preprocessed and analyzed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB (version R2023b; The MathWorks, Inc.) (see detailed Supplemental Methods). Demographic and clinical data were analyzed using SPSS version 29 (IBM Corp.). Normally distributed data were compared using *t* tests; non-normally distributed data were compared using Mann-Whitney *U* tests; and categorical data were compared using χ^2 tests. Repeated-measures analysis of variance was used to compare group differences in wanting and liking ratings, with the $HDRS_{17}$ score as a covariate. Imaging data were analyzed with an event-related random effects design. Effects were modeled for CS and US conditions (time points 3 and 6, respectively). Parametric modulations were included by entering TD PE values for both CS and US. Functional connectivity between the habenula and VTA during aversive learning was investigated with a generalized psychophysiological interaction (gPPI) analysis with the bilateral habenula as the seed region (33). Group and sex differences were examined with 2-sample *t* tests. Associations between 1) the habenula TD signal and residual symptoms ($HDRS_{17}$) and 2) habenula-VTA gPPI connectivity and residual symptoms were evaluated with 2 separate multiple regression analyses. Given the habenula's proximity to the cerebral cisterns, we repeated the analysis using the third ventricle as the ROI to determine whether the effects observed in the habenula were specific to this region rather than being due to spurious signal in the cerebrospinal fluid. More detailed information about the statistical analysis is provided in the Supplement.

RESULTS

Sample Characteristics

Table 1 shows characteristics of the included participants. Residual severity scores ($HDRS_{17}$; 3.5 vs. 1, $p < .001$) were significantly higher in patients than in control participants. At the time of inclusion, all participants had an $HDRS_{17}$ score of ≤7. However, at the time of MRI acquisition, 5 participants had an $HDRS_{17}$ score between 9 and 16, indicating worsening of symptoms. These participants, who did not meet the criteria for a new depressive episode, had not restarted their antidepressants before the scan. Of these 5 participants, 4 experienced a relapse later during the study period.

Table 1. Demographic and Clinical Characteristics of Study Participants

Characteristic	rrMDD, *n* = 36	Healthy Control, *n* = 27
Age, Years	54 [36–65]	50 [36–63]
Sex		
Female	26 (72.2%)	19 (70.4%)
Male	10 (27.6%)	8 (29.6%)
Education Levels		
Low	2 (5.6%)	0 (0.0%)
Middle	14 (38.9%)	13 (48.1%)
High	20 (55.6%)	14 (51.9%)
IQ	108 (8.9)	105 (9.9)
HDRS Intake, Median (IQR)	3 (4)	0 (1)
HDRS MRI, Median (IQR)	3.5 (4)	1 (2)
Number of Lifetime Episodes	9.2 (11.7)	–
Age of Onset, Years	24.7 (10.9)	–

Values are presented as *n* (%), mean (SD), or mean [range] unless otherwise specified.

HDRS, Hamilton Depression Rating Scale; IQR, interquartile range; MRI, magnetic resonance imaging; rrMDD, remitted recurrent major depressive disorder.

Behavioral Results

Participants rated the aversive stimulus in terms of wanting and liking. The mean wanting score was −0.94 (SD = 0.76), while the mean liking score was −0.99 (SD = 0.84) on a scale ranging from −2 to 2. These negative scores indicate that participants perceived the stimulus as aversive. No main effect of time or group was demonstrated for wanting or liking ratings. Moreover, no significant group × time interactions were observed (Figure 3).

fMRI Results

We observed a main effect of TD-related activation in the bilateral habenula (Table 2 and Figure 4A), showing that our reinforcement learning task was capable of eliciting habenula activation. Furthermore, we demonstrated increased TD error–related activity (CS × TD contrast) in patients with rrMDD compared with HCs in the left and right habenula ($z = 2.89$ and 2.79; familywise error small volume–corrected p [$p_{FWE\text{-}SVC}$] = .012 and .016, respectively). Mismatches between aversive expectations and current states with respect to aversive outcomes (as modeled by the TD model) at the presentation of the CS led to higher BOLD responses in the habenula for patients with rrMDD than for HCs (Table 3 and Figure 4B). Importantly, this effect remained significant for the right habenula even after patients with an $HDRS_{17}$ score >7 were excluded ($z = 2.41$; $p_{FWE\text{-}SVC} = .044$). No sex differences were found in habenula response to aversive stimuli ($t_{61} = 0.247$, $p = .805$). No group differences were found in the CS × TD + US × TD and US × TD contrast. For all contrasts, no significant group differences were observed when voxels from the adjacent third ventricle were analyzed (see Table S1).

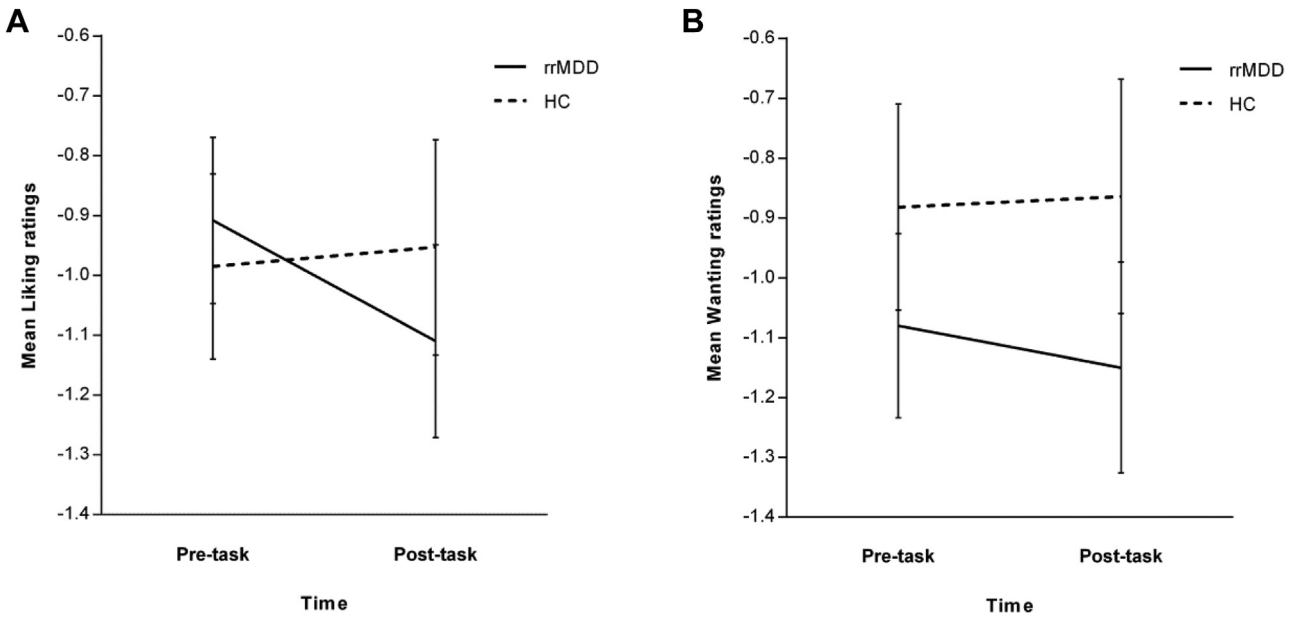

Figure 3. Liking and wanting ratings. **(A)** Liking ratings (estimated marginal means ± 1 SEM [adjusted for any other variables in the model]): no significant main effect of group ($F_{1,57}$ = 0.03, p = .854), no significant main effect of time ($F_{1,57}$ = 0.42, p = .519), and no significant group × time interaction ($F_{1,57}$ = 1.21, p = .277). **(B)** Wanting ratings (estimated marginal means ± 1 SEM [adjusted for any other variables in the model]): no significant main effect of group ($F_{1,57}$ = 1.06, p = .307), no significant main effect of time ($F_{1,57}$ = 2.34, p = .131), and no significant group × time interaction ($F_{1,57}$ = 0.16, p = .691). HC, healthy control participant; rrMDD, remitted recurrent major depressive disorder.

The gPPI analyses revealed main habenula-VTA connectivity regardless of task ($z > 8$, $p < .001$), but there were no differences between groups. Moreover, we observed main habenula-VTA connectivity during the task regardless of TD modulation ($z > 8$, $p < .001$), but again there were no group differences. During aversive learning, which was modulated by TD errors (CS × TD contrast), we found main habenula-VTA connectivity (z = 2.78, p = .003) that differed between the groups: Patients with rrMDD exhibited decreased functional connectivity as a function of TD (the degree of expectation of punishment during CS presentation) between the habenula and the VTA compared with HCs (z = 3.73, p = .002) (see Tables 2 and 3; Figure 4C). This difference remained significant even after we excluded patients with an $HDRS_{17}$ score >7 (z = 3.94, p = .001). During aversive learning, which was modulated by TD errors (CS × TD contrast), we found no significant sex differences in habenula-VTA connectivity (t_{61} = 0.136 p = .893). No group differences were observed in functional connectivity between the third ventricle and the VTA during TD-modulated aversive learning (see Table S1).

Association Between fMRI Results and Residual Symptoms

The regression model with $HDRS_{17}$ scores, group, and their interaction as independent variables and aversive learning signals in the habenula as the dependent variable showed a significant group × $HDRS_{17}$ interaction (t_{59} = 2.37, p = .021) (Figure 5), which explained 21.8% of the variance of the total model ($F_{3,62}$ = 5.478, p = .002). No main effect of group or $HDRS_{17}$ was observed. In the rrMDD group, higher

Table 2. Main Effects (Activation and Connectivity)

Activation and Connectivity	Contrast	Location	MNI Coordinates: x, y, z	z	Significance
TD Activation					
Total TD signal, CS × TD + US × TD	rrMDD + HCs	Habenula	3, −21, 3	1.81	.035
			−3, −24, 3	1.76	.039
Cue presentation, CS × TD	rrMDD + HCs	Habenula	3, −21, 3	2.53	.006
			−3, −24, 3	2.59	.005
Outcome delivery, US × TD	rrMDD + HCs	–	No significant habenula cluster		
gPPI: Habenula Seed					
Main connectivity regardless of task	rrMDD + HCs	VTA	0, −21, 0	>8	<.001
Main connectivity task without TD	rrMDD + HCs	VTA	−6, −21, −3	>8	<.001
Main connectivity task with TD	rrMDD + HCs	VTA	3, −21, −6	2.78	.003

CS, conditioned stimuli; gPPI, generalized psychophysiological interaction; HC, healthy control participant; MNI, Montreal Neurological Institute; rrMDD, remitted recurrent major depressive disorder; TD, temporal difference; US, unconditioned stimuli; VTA, ventral tegmental area.

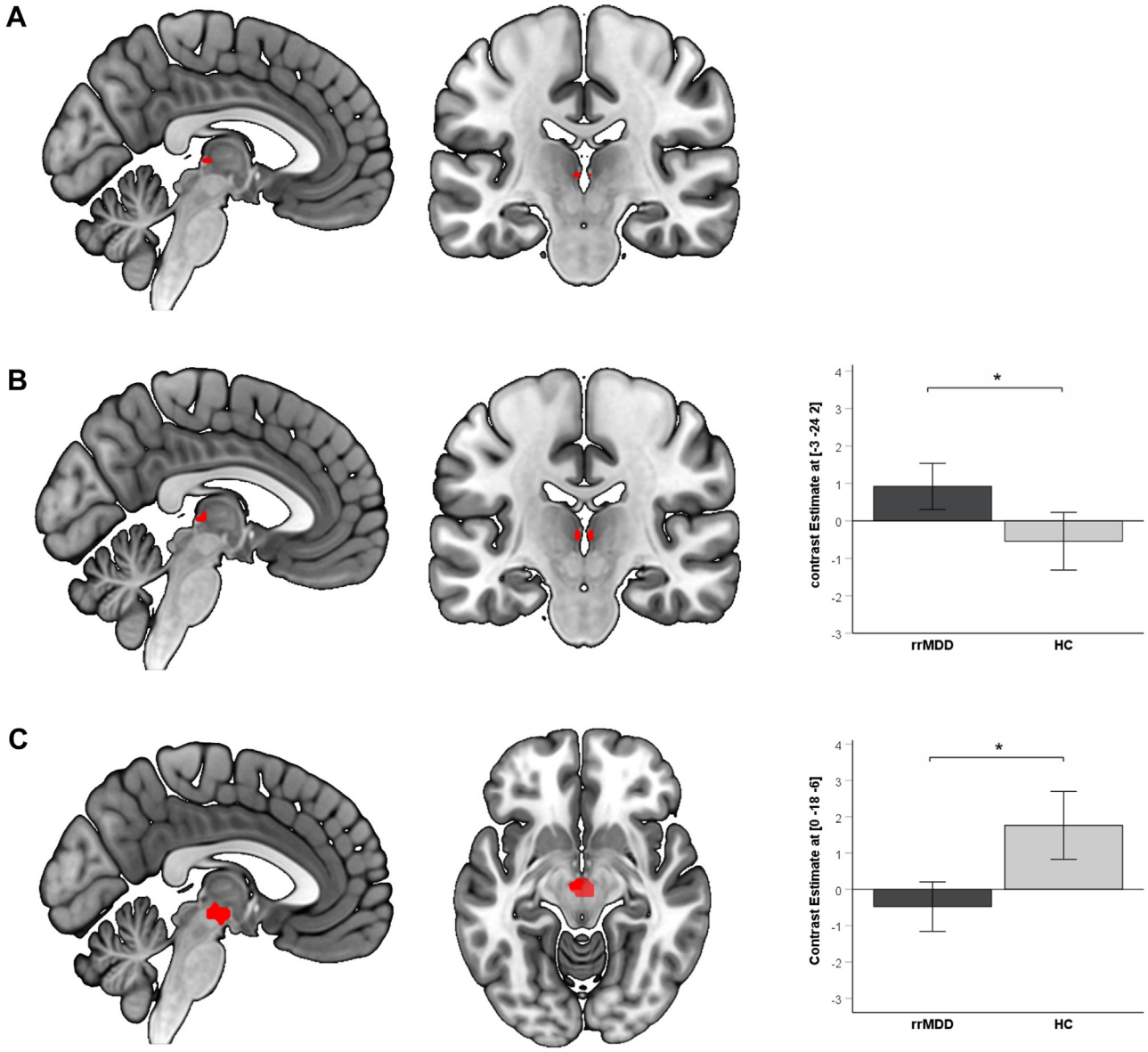

Figure 4. **(A)** Main effect of task. Temporal difference (TD) error–related habenula activation in the CS × TD + US × TD contrast (left habenula: $z = 1.83$, $p = .034$; right habenula: $z = 1.71$, $p = .044$, uncorrected to display the extend of the signal). **(B)** TD error–related activation comparing patients with remitted recurrent major depressive disorder (rrMDD) vs. healthy control participants (HCs). Patients with rrMDD show more TD-related activation (CS × TD contrast) in the bilateral habenula than HCs (left habenula: $z = 2.89$, $p = .012$; right habenula: $z = 2.79$, $p = .016$ familywise error [FWE] corrected on peak level, small volume corrected). **(C)** Generalized psychophysiological interaction results with the habenula as seed. Patients with rrMDD showed decreased functional connectivity between the habenula and the ventral tegmental area compared with HCs during aversive learning ($z = 3.73$, $p = .002$, FWE corrected on peak level, small volume corrected). *$p < .05$. CS, conditioned stimulus; US, unconditioned stimulus.

residual severity, as measured by the $HDRS_{17}$, positively correlated with more aversive learning signals in the habenula ($r = 0.407$, $p = .014$). After exclusion of 2 outliers with moderate depressive symptoms when scanned, these effects in the rrMDD group were no longer significant ($r = 0.210$, $p = .233$).

For habenula-VTA connectivity, the regression model with $HDRS_{17}$ scores, group, and group × $HDRS_{17}$ interaction explained 23% of the variance of the total model ($F_{3,62} = 5.78$, $p = .002$). A main effect of group was observed ($t_{59} = -3.60$, $p = .001$). No main effect of $HDRS_{17}$ scores and no significant group × $HDRS_{17}$ interaction were found. In the rrMDD group, higher residual severity, as measured by $HDRS_{17}$ scores, was not significantly correlated with habenula-VTA connectivity. Exclusion of the 2 outliers did not change these results.

Table 3. Between-Group Differences in Activation and Connectivity

Activation and Connectivity	Contrast	Location	MNI Coordinates: x, y, z	*z*	Significance
TD Signal					
Total TD signal, CS × TD + US × TD	rrMDD > HC	Habenula	−3, −24, 3	2.29	.059
			3, −21, 3	2.19	.081
	HC > rrMDD	No clusters survived threshold			
CS × TD	rrMDD > HC	Habenula	−3, −24, 3	2.89	.012
			3, −21, 3	2.79	.016
	HC > rrMDD	No clusters survived threshold			
US × TD	rrMDD > HC	No clusters survived threshold			
	HC > rrMDD	No clusters survived threshold			
gPPI					
Habenula seed	HC > rrMDD	VTA	0, −18, −6	3.73	.002

CS, conditioned stimuli; gPPI, generalized psychophysiological interaction; HC, healthy control participant; MNI, Montreal Neurological Institute; rrMDD, remitted recurrent major depressive disorder; TD, temporal difference; US, unconditioned stimuli; VTA, ventral tegmental area.

DISCUSSION

Using a classical conditioning task, we investigated the neural response of the habenula during aversive learning in medication-free patients with rrMDD compared with HCs. We found that, relative to HCs, patients with rrMDD showed significantly increased TD-related activation in the bilateral habenula, with a positive correlation between signal magnitude and the severity of residual symptoms. Furthermore, patients with rrMDD exhibited decreased functional connectivity between the habenula and the VTA, based on the TD-related activations during CS presentation.

The analysis of behavioral responses revealed no group differences in subjective wanting or liking processing of aversive stimuli. A similar lack of group differences has been described before in a sample of patients with remitted depression (34). This indicates that the increased BOLD responses demonstrated here were not driven by subjective reportable differences in how aversive stimuli were anticipated and experienced.

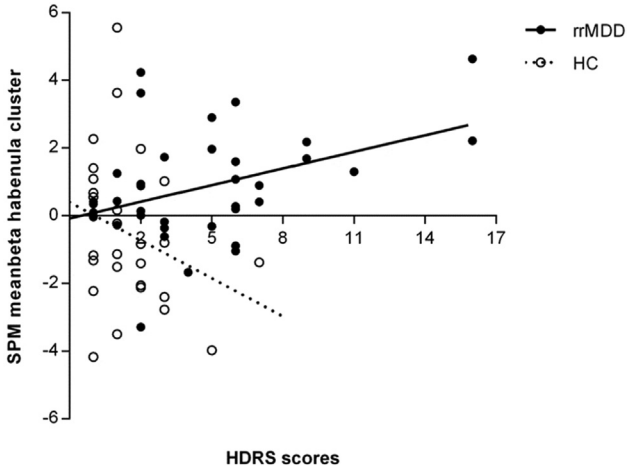

Figure 5. Association of aversive prediction error signal in the habenula with residual Hamilton Depression Rating Scale (HDRS) ratings. Significant group × HDRS interaction (t_{59} = 2.37, p = .021) and significant positive habenula-HDRS correlation in the remitted recurrent MDD (rrMDD) group (r = 0.407, p = .014). HC, healthy control participant.

Our findings of aberrant response of the habenula in our patient group could be interpreted in the context of studies that have used animal models for depression. These studies indicate hyperactivity of the habenula (16,17) and suggest that the habenula plays a central role in behavioral responses to punishment (35). It has been hypothesized that overactivity of the habenula may lead to increased sensitivity to aversive events (16). In our study, this increased habenula activity during aversive learning was specifically apparent during CS processing. This may reflect an overly active coupling between environmental cues and punishment, being present when in remission from MDD, which may be important for recurrence vulnerability. However, whether this represents a trait-like characteristic needs to be determined by replication in larger longitudinal samples.

While our research is partially consistent with findings from animal studies, previous investigations of humans have shown blunted habenula activation (9,10) or even decreased habenula activation in patients with MDD (11). An important difference is that we investigated patients with remitted MDD instead of patients with current depression. A distinction in the interplay between tonic and phasic activity of the habenula may exist between patients with remitted depression and patients who are currently depressed. It could be hypothesized that the habenula shows heightened tonic activation in individuals with current MDD, potentially resulting in diminished relative effects of phasic responsivity (reactive bursting in response to stimuli). This hypothesis of heighted tonic activation gains support from a positron emission tomography (PET) study that revealed a correlation between depressive symptoms and increased metabolism of the habenula following tryptophan depletion (36). Notably, this study was conducted with 8 male patients with remitted MDD, most of whom were on antidepressant treatment. Therefore, this hypothesis warrants additional investigation by combining PET with fMRI.

A key distinction between our study and previous research on habenula activity during aversive learning is the type of reinforcer used. While 2 earlier studies relied on secondary reinforcers (e.g., money), we used a primary reinforcer (aversive liquid), which has innate biological significance. In

contrast, secondary reinforcers gain value through learned associations. Lawson *et al.* (11) also used a primary reinforcer (electric shock) and found that habenula activation in individuals with MDD decreased as CS became more strongly associated with the shock (11). This finding highlights how habenula responses to primary aversive reinforcers may not only vary depending on the specific reinforcer used but may also be modulated by differences in clinical populations (current vs. remitted MDD). Ulrich *et al.* (37) examined negative PE activity for sex, food, and money rewards and showed that the anterior insula responded to all reinforcers, whereas the VTA responded to sex and money, and the habenula responded specifically to sex (37). Therefore, even among primary reinforcers, responses vary according to the region being studied. Thus, caution is warranted when comparing our study with previous research, because differences in reinforcer type might have influenced neural activation patterns.

Differences in results could also be attributed to the complexity of interpreting the BOLD signal. In fMRI, the measured BOLD signal may indicate the activation of both excitatory and inhibitory synapses (38). Taking these considerations into account, our findings nevertheless add to the evidence that aberrant habenula activity during aversive learning is associated with depression and extend the evidence by showing that aberrant habenula functioning can persist during remission.

The positive association between TD aversive learning activation in the habenula and residual symptom severity ratings in patients with rrMDD is consistent with Liu *et al.* (10), who found greater habenula activation in patients with more severe depression. This association may result from reduced serotonergic input from the raphe nuclei (39). Reduced serotonergic input may elevate the activity of the habenula, which in turn inhibits the dopaminergic VTA, which together mediate depressive symptoms (15,39). However, in our study, the association between habenula activity and residual illness severity was no longer significant after the exclusion of 2 outliers, which either reduces the contrast of the observations or suggests a possible nonlinear association and therefore should be interpreted with caution.

Furthermore, we found decreased functional connectivity as a function of the TD signal between the habenula and the VTA in patients with rrMDD compared with HCs. In healthy participants, functional coupling between the habenula and VTA during aversive stimulation has been established (40,41). Because of the (indirect) inhibitory influence of the habenula on the VTA, aberrant activity of the habenula in combination with decreased functional connectivity may influence this functional coupling and could result in a net decoupling of the VTA from the habenula. Our connectivity results are surprising when compared with Kumar *et al.* (26), who found no differences between patients with MDD and HCs in VTA-habenula connectivity during loss trials of a monetary instrumental learning task. One possibility is that Kumar *et al.* (26) investigated functional connectivity between the habenula and the VTA without TD modulation whereas we observed group differences in functional connectivity as a function of TD PE (i.e., the difference between expectation of punishment and the actual situation). When exploring our results without TD modulation, we also found no group differences. This suggests that the TD modulation may give a more accurate representation of aversion-related functional connectivity between the habenula and VTA. Importantly, differences between paradigms (e.g., primary vs. secondary reinforcement) may also account for these differences, because animal studies suggest that the habenula and VTA predominantly show firing with primary reinforcements (42). Therefore, our task may be more sensitive in mapping connectivity between the habenula and VTA as opposed to the paradigm with secondary reinforcements (i.e., monetary) (26).

Some limitations of the current study are worth mentioning. First, fMRI resolution is limited for small structures such as the VTA and the habenula. This complicates the distinction of the lateral from the medial habenula. Nonetheless, by eliminating physiological noise with DRIFTER (43) in conjunction with a priori computational modeling specific for aversive learning, we interpret our findings as representing aversive learning–related differences between groups and thereby valuable in contributing to finding the pathophysiology that underlies recurrence in depression. Furthermore, the additional analysis with the third ventricle as ROI substantiates that our main results are not due to spurious signal in adjacent cerebrospinal fluid. Second, the exclusion of participants due to invalid physiological noise data is another limitation of this study. While including these participants could increase sample size, we chose a priori to correct for physiological noise, as is commonly done in similar studies, because it significantly impacts oxygenation-sensitive neuroimaging, particularly in small regions such as the habenula and VTA (11,30,44,45). Third, we used gPPI to investigate functional connectivity between the habenula and VTA. PPI measures the temporal correlation between remote neuronal activity but does not specify the direction of influence between brain regions. Effective connectivity could measure the influence that one neural system exerts over another (46). Fourth, a potential limitation of our study is that we predefined the learning rates rather than fitting them to individual data because the design of our study did not allow for the latter. While this predefined approach has been used in previous research, it may not provide the optimal parameters for this specific dataset. However, previous work suggests that variations in learning rate may have a limited impact on neural results in model-based fMRI, implying that precise model fitting is not always critical (28). Finally, the high proportion of highly educated participants in our sample, although not significantly different between groups, potentially complicates the generalizability of our findings.

Conclusions

We demonstrated aberrant habenula activity and decreased habenula-VTA functional connectivity during aversive learning in patients with rrMDD compared with HCs. Collectively, these findings reveal insight into the involvement of aberrant aversive learning habenula functioning during remission. Future studies should examine—within the same participants—how different reinforcer types influence neural activation patterns, because variations in primary and secondary reinforcers may contribute to differences in aversive learning and associated neural processes. It is also important to determine whether habenula dysfunction represents a state-dependent or trait-like

characteristic. Longitudinal research could investigate whether habenula dysfunction fluctuates with changes in clinical status, such as relapse versus remission in patients with rrMDD, but also whether it can be used as a predictor for relapse.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by a dedicated grant from the Dutch Brain Foundation (Grant No. 2009(2)-72 [to HGR]). HGR is supported by NWO/ZonMw VENI (Grant No. 016.126.059). The funder did not participate in designing or conducting the study.

Presented as a poster at the 36th European College of Neuropsychopharmacology Congress, October 2023, Barcelona, Spain. Parts of this data have also been published online as part of HG's thesis.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

HGR reports speaking fees from Janssen and Lundbeck and an educational grant from Janssen, all outside the current work. All other authors report no biomedical financial interests or potential conflicts of interest.

Dutch Clinical Trials Register: DELTA-neuroimaging; https://www.onderzoekmetmensen.nl/nl/trial/27041; NTR3768.

ARTICLE INFORMATION

From the Radboud University Medical Center, Department of Psychiatry, Nijmegen, the Netherlands (JMdK-S, PFPvE, DEMG, HGR); Donders Institute for Brain Cognition and Behavior, Nijmegen, the Netherlands (JMdK-S, PFPvE, DEMG, HGR); Martini Ziekenhuis, Waardegedreven Verbeteren, Groningen, the Netherlands (HG); Amsterdam University Medical Center, Department of Psychiatry, Amsterdam, the Netherlands (RJTM); Delft University of Technology, Department of Technology, Policy and Management, Delft, the Netherlands (CAF); Amsterdam University Medical Center, Department of Radiology and Nuclear Medicine, Amsterdam, the Netherlands (PFCG); and University Medical Center Groningen, Department of Biomedical Sciences of Cells & Systems, Cognitive Neuroscience Center, Groningen, the Netherlands (J-BCM).

JMdK-S and HG are joint first authors.

Address correspondence to Jessica M. de Klerk-Sluis, M.D., at Jessica.deKlerk-Sluis@radboudumc.nl.

Received Dec 5, 2024; revised Apr 8, 2025; accepted Apr 9, 2025.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.bpsc.2025.04.006.

REFERENCES

1. Bockting CLH, Spinhoven P, Wouters LF, Koeter MWJ, Schene AH, DELTA Study Group (2009): Group Long-term effects of preventive cognitive therapy in recurrent depression: A 5.5-year follow-up study. J Clin Psychiatry 70:1621–1628.
2. Greenberg PE, Fournier AA, Sisitsky T, Simes M, Berman R, Koenigsberg SH, Kessler RC (2021): The economic burden of adults with major depressive disorder in the United States (2010 and 2018). Pharmacoeconomics 39:653–665.
3. Eshel N, Roiser JP (2010): Reward and punishment processing in depression. Biol Psychiatry 68:118–124.
4. Jean-Richard-Dit-Bressel P, Killcross S, McNally GP (2018): Behavioral and neurobiological mechanisms of punishment: Implications for psychiatric disorders. Neuropsychopharmacology 43:1639–1650.
5. Santesso DL, Steele KT, Bogdan R, Holmes AJ, Deveney CM, Meites TM, Pizzagalli DA (2008): Enhanced negative feedback responses in remitted depression. NeuroReport 19:1045–1048.
6. Rothkirch M, Tonn J, Köhler S, Sterzer P (2017): Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. Brain 140:1147–1157.
7. Garrison J, Erdeniz B, Done J (2013): Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310.
8. Sutton RS (1988): Learning to predict by the methods of temporal differences. Mach Learn 3:9–44.
9. Furman DJ, Gotlib IH (2016): Habenula responses to potential and actual loss in major depression: Preliminary evidence for lateralized dysfunction. Soc Cogn Affect Neurosci 11:843–851.
10. Liu WH, Valton V, Wang LZ, Zhu YH, Roiser JP (2017): Association between habenula dysfunction and motivational symptoms in unmedicated major depressive disorder. Soc Cogn Affect Neurosci 12:1520–1533.
11. Lawson RP, Nord CL, Seymour B, Thomas DL, Dayan P, Pilling S, Roiser JP (2017): Disrupted habenula function in major depression. Mol Psychiatry 22:202–208.
12. Gold PW, Kadriu B (2019): A major role for the lateral habenula in depressive illness: Physiologic and molecular mechanisms. Front Psychiatry 10:320.
13. Jin IB, Kim NH, Han JS (2024): Emerging role of the lateral habenula in conditioned inhibition and depression. Behav Neurosci 138:212–220.
14. Cameron S, Weston-Green K, Newell KA (2024): The disappointment centre of the brain gets exciting: A systematic review of habenula dysfunction in depression. Transl Psychiatry 14:499.
15. Zhang H, Li K, Chen HS, Gao SQ, Xia ZX, Zhang JT, *et al.* (2018): Dorsal raphe projection inhibits the excitatory inputs on lateral habenula and alleviates depressive behaviors in rats. Brain Struct Funct 223:2243–2258.
16. Proulx CD, Hikosaka O, Malinow R (2014): Reward processing by the lateral habenula in normal and depressive behaviors. Nat Neurosci 17:1146–1152.
17. Yang Y, Wang H, Hu J, Hu H (2018): Lateral habenula in the pathophysiology of depression. Curr Opin Neurobiol 48:90–96.
18. Li B, Piriz J, Mirrione M, Chung C, Proulx CD, Schulz D, *et al.* (2011): Synaptic potentiation onto habenula neurons in the learned helplessness model of depression. Nature 470:535–539.
19. Yang Y, Cui Y, Sang K, Dong Y, Ni Z, Ma S, Hu H (2018): Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. Nature 554:317–322.
20. Sartorius A, Kiening KL, Kirsch P, von Gall CC, Haberkorn U, Unterberg AW, *et al.* (2010): Remission of major depression under deep brain stimulation of the lateral habenula in a therapy-refractory patient. Biol Psychiatry 67:e9–e11.
21. Chen C, Takahashi T, Nakagawa S, Inoue T, Kusumi I (2015): Reinforcement learning in depression: A review of computational research. Neurosci Biobehav Rev 55:247–267.
22. Mocking RJT, Figueroa CA, Rive MM, Geugies H, Servaas MN, Assies J, *et al.* (2016): Vulnerability for new episodes in recurrent major depressive disorder: Protocol for the longitudinal DELTA-neuroimaging cohort study. BMJ Open 6:e009510.
23. Hamilton M (1967): Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278–296.
24. Kumar P, Waiter G, Ahearn T, Milders M, Reid I, Steele JD (2008): Abnormal temporal difference reward-learning signals in major depression. Brain 131:2084–2093.
25. Dayan P, Abbott F (2001): Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. Cambridge, MA: The MIT Press.
26. Kumar P, Goer F, Murray L, Dillon DG, Beltzer ML, Cohen AL, *et al.* (2018): Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. Neuropsychopharmacology 43:1581–1588.
27. Daw ND (2011): Trial-by-trial data analysis using computational models. In: Delgado MR, Phelps EA, Robbins TW, editors. (2011), Decision Making, Affect, and Learning. Attention and Performance, vol. XXIII:New York: Oxford University Press Incorporated, 3–38.
28. Wilson RC, Niv Y (2015): Is model fitting necessary for model-based fMRI? PLOS Comput Biol 11:e1004237.
29. Geugies H, Mocking RJT, Figueroa CA, Groot PFC, Marsman JC, Servaas MN, *et al.* (2019): Impaired reward-related learning signals in remitted unmedicated patients with recurrent depression. Brain 142:2510–2522.

30. Lawson RP, Seymour B, Loh E, Lutti A, Dolan RJ, Dayan P, *et al.* (2014): The habenula encodes negative motivational value associated with primary punishment in humans. Proc Natl Acad Sci USA 111:11858–11863.
31. Liu TT, Frank LR, Wong EC, Buxton RB (2001): Detection power, estimation efficiency, and predictability in event-related fMRI. Neuroimage 13:759–773.
32. Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M, *et al.* (2011): Expected value and prediction error abnormalities in depression and schizophrenia. Brain 134:1751–1764.
33. McLaren DG, Ries ML, Xu G, Johnson SC (2012): A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. Neuroimage 61:1277–1286.
34. McCabe C, Cowen PJ, Harmer CJ (2009): Neural representation of reward in recovered depressed patients. Psychopharmacology 205:667–677.
35. Wilcox KS, Christoph GR, Double BA, Leonzio RJ (1986): Kainate and electrolytic lesions of the lateral habenula: Effect on avoidance responses. Physiol Behav 36:413–417.
36. Morris JS, Smith KA, Cowen PJ, Friston KJ, Dolan RJ (1999): Covariation of activity in habenula and dorsal raphé nuclei following tryptophan depletion. Neuroimage 10:163–172.
37. Ulrich M, Rüger A, Durner V, Grön G, Graf H (2023): Reward is not reward: Differential impacts of primary and secondary rewards on expectation, outcome, and prediction error in the human brain's reward processing regions. Neuroimage 283:120440.
38. Logothetis NK (2008): What we can do and what we cannot do with fMRI. Nature 453:869–878.
39. Zhao H, Zhang BL, Yang SJ, Rusak B (2015): The role of lateral habenula-dorsal raphe nucleus circuits in higher brain functions and psychiatric illness. Behav Brain Res 277:89–98.
40. Ide JS, Li CSR (2011): Error-related functional connectivity of the habenula in humans. Front Hum Neurosci 5:25.
41. Hennigan K, D'Ardenne K, McClure SM (2015): Distinct midbrain and habenula pathways are involved in processing aversive events in humans. J Neurosci 35:198–208.
42. Matsumoto M, Hikosaka O (2007): Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447:1111–1115.
43. Särkkä S, Solin A, Nummenmaa A, Vehtari A, Auranen T, Vanni S, Lin FH (2012): Dynamic retrospective filtering of physiological noise in BOLD fMRI: DRIFTER. Neuroimage 60:1517–1527.
44. D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008): BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319:1264–1267.
45. Krüger G, Glover GH (2001): Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn Reson Med 46:631–637.
46. Friston KJ (2011): Functional and effective connectivity: A review. Brain Connect 1:13–36.