Impaired reward-related learning signals in remitted unmedicated patients with recurrent depression

Hanneke Geugies,1,2 Roel J.T. Mocking,3 Caroline A. Figueroa,3,4 Paul F.C. Groot,5 Jan-Bernard C. Marsman,2 Michelle N. Servaas,2 J. Douglas Steele,6 Aart H. Schene7,8 and Henricus G. Ruhe1,3,4,8

One of the core symptoms of major depressive disorder is anhedonia, an inability to experience pleasure. In patients with major depressive disorder, a dysfunctional reward-system may exist, with blunted temporal difference reward-related learning signals in the ventral striatum and increased temporal difference-related (dopaminergic) activation in the ventral tegmental area. Anhedonia often remains as residual symptom during remission; however, it remains largely unknown whether the abovementioned reward systems are still dysfunctional when patients are in remission. We used a Pavlovian classical conditioning functional MRI task to explore the relationship between anhedonia and the temporal difference-related response of the ventral tegmental area and ventral striatum in medication-free remitted recurrent depression patients \( (n = 36) \) versus healthy control subjects \( (n = 27) \). Computational modelling was used to obtain the expected temporal difference errors during this task. Patients, compared to healthy controls, showed significantly increased temporal difference reward learning activation in the ventral tegmental area \( (P_{FWESVC} = 0.028) \). No differences were observed between groups for ventral striatum activity. A group \( \times \) anhedonia interaction \( [t(57) = -2.29, P = 0.026] \) indicated that in patients, higher anhedonia was associated with lower temporal difference activation in the ventral tegmental area, while in healthy controls higher anhedonia was associated with higher ventral tegmental area activation. These findings suggest impaired reward-related learning signals in the ventral tegmental area during remission in patients with depression. This merits further investigation to identify impaired reward-related learning as an endophenotype for recurrent depression. Moreover, the inverse association between reinforcement learning and anhedonia in patients implies an additional disturbing influence of anhedonia on reward-related learning or vice versa, suggesting that the level of anhedonia should be considered in behavioural treatments.
Tempral difference learning signals in recurrent depression

Introduction

Major depressive disorder (MDD) is a highly prevalent and disabling disease (Mathers and Loncar, 2006). Although treatment of a depressive episode can induce remission of symptoms, depressive episodes unfortunately tend to recur after a period of recovery (Frank et al., 1991). The incidence of recurrences varies (depending on the population and setting) but may reach 80% within 5 years (Bockting et al., 2009). Therefore, recurrence is a major contributor to the immense (in)direct annual costs of MDD (estimated >€113 billion in Europe) (Gustavsson et al., 2011), which necessitates prevention of recurrence and knowledge of underlying aetiopathogenetic mechanisms.

An inability to experience pleasure/reward (anhedonia) is one of the core symptoms of depression (Ebmeier et al., 2006) and often persists as a residual symptom after remission (Conradi et al., 2011). The ability to experience reward appears important in providing resilience against recurrence. Positive emotional responses decrease stress-sensitivity (Wichers et al., 2007), and predict recovery during antidepressant treatment (Wichers et al., 2009). Furthermore, pleasure also has an important motivational function; it reinforces behaviour that leads to (potentially) pleasurable events (conditioning) (Pavlov, 1927). Patients with MDD often report either difficulties in experiencing normally positive events as pleasurable (i.e. consummatory anhedonia or ‘liking’) or deficits in motivation to pursue rewards (i.e. motivational anhedonia or ‘wanting’) (Treadway and Zald, 2011). Furthermore, patients with MDD have difficulties in learning new behaviours that might improve their mood or keep them well (Vrieze et al., 2013).

Wanting, liking and learning have been identified as three important dissociable components of reward (Berridge et al., 2009), where especially wanting and learning have been linked to dopaminergic neurotransmission in the reward-network consisting of the ventral striatum (Knutson et al., 2001; Schott et al., 2008) and ventral tegmental area (D’Ardenne et al., 2008; Kumar et al., 2008; Schott et al., 2008). In the reward circuitry, the ventral tegmental area projects to the ventral striatum and receives projections from the habenula, which is involved in regulating the intensity of reward-seeking and distress-avoiding behaviour (Loonen and Ivanova, 2017).

Previous studies have shown that reward learning stimuli evoke short phasic firing patterns of dopaminergic neurons (Schultz, 1998; Tobler et al., 2005), resembling temporal difference prediction errors (Schultz et al., 1997; Kumar et al., 2008). Temporal difference prediction errors are important for making a predictive association between stimuli and outcomes when stimuli are repeated and learned. Over time, dopaminergic neurons will predict a response as a result of previous associations between a stimulus and its rewarding value (classical conditioning/reinforcement learning). Briefly, before learning, delivery of an unexpected reward is followed by phasic dopamine activation. When the association between stimulus and reward has been consolidated, dopaminergic firing is activated at the presentation of the stimulus (cue), while firing to the reward itself is reduced when delivered as expected. However, when a learned cue is not followed by an expected reward, this results in a decrease in dopaminergic firing (below baseline), representing negative prediction errors.

Dysfunctions in anticipatory and consummatory reward processes in MDD have been investigated (Knutson et al., 2008; Pizzagalli et al., 2009; Smoski et al., 2009), as well as temporal difference reward-related learning in depressed patients versus control subjects (Kumar et al., 2008). Kumar and colleagues identified increased activation of dopaminergic neurons in the ventral tegmental area when thirsty patients with MDD were learning associations between a stimulus (picture) and a reward (water delivery) (Kumar et al., 2008). Furthermore, the ventral striatum has been repeatedly reported to be hypoactive in MDD both in reinforcement-learning as in other reward processing paradigms (Kumar et al., 2008; Pizzagalli et al., 2009; Gradin et al., 2011; Robinson et al., 2012; Hall et al., 2014).

Although evidence for a dysfunctional reward system in depressed patients is established (Martin-Soelch, 2009), there is still very little understanding whether these reward systems remain dysfunctional when patients are in remission. Previous studies conducted in subjects at risk for depression and with subthreshold depression have demonstrated that abnormalities in processing of wanting and liking aspects of reward may be a trait marker for MDD (McCabe et al., 2009, 2012; Stringaris et al., 2015; McCabe, 2016; Pan et al., 2017). However, it remains largely unknown whether a dysfunction in processing of reward-related learning represents a trait rather than a
state-dependent abnormality, which may be of importance with regard to vulnerability for recurrence. Furthermore, little is known about the association between persistent anhedonia and deficits of reward processing in remitted patients (Dunlop and Nemeroff, 2007). We therefore quantified the response of the dopamine reward system (i.e. ventral striatum and ventral tegmental area) during a classical conditioning functional MRI task in medication-free patients with remitted recurrent depression (rrMDD), who were at high risk of recurrence (Mocking et al., 2016). In addition, we hypothesized a link between abnormalities in the reward system and anhedonia levels. Based on earlier work in depressed patients during classical conditioning (Kumar et al., 2008), we hypothesized decreased ventral striatum activation and increased ventral tegmental area activation in response to temporal difference reward-related learning in rrMDD versus controls, with positive associations of these abnormalities with anhedonia.

**Materials and methods**

**Participants**

As part of a larger neuroimaging study investigating vulnerability for recurrence in MDD (Mocking et al., 2016), participants were recruited by advertisements and through previous clinical treatment and/or previous studies. In particular, patients aged 35–65 with a known recurrent depressive disorder, currently in stable remission without medication, were identified and approached for this study. Matched healthy control subjects were recruited via advertisements. We obtained permission from the local ethics committee and written informed consent from all participants (Mocking et al., 2016). Dimensional assessment of illness severity was obtained by an observer-rated Hamilton Depression Rating Scale (HDRS17) (Hamilton, 1967), and a self-rated Snath Hamilton Anhedonia and Pleasure Scale (SHAPS) (Snath et al., 1995). Sixty-two patients with MDD were scanned who satisfied the following criteria: (i) presence of a recurrent depression defined as ≥2 depressive episodes according to the structured interview for DSM-IV (SCID); (ii) stable remission defined as a HDRS17 ≤ 7 for at least eight subsequent weeks; and (iii) aged between 35–65. We scanned 41 healthy controls that were matched on the basis of age, sex and years of education. All participants were without any medications for >4 weeks. Exclusion criteria were: (i) a current diagnosis of alcohol or drug dependence; (ii) psychotic or bipolar disorder; (iii) primary anxiety disorder; (iv) MRI participation contraindications such as implanted metal; (v) electroconvulsive therapy within 2 months before scanning; and (vi) a history of head trauma or neurological disease. Healthy controls were excluded if they had personal (SCID) or first degree relatives with a psychiatric disorder.

**Task**

A Pavlovian classical conditioning task was used specifically to assess reward learning during passive observation (Kumar et al., 2008) instead of an instrumental design that would have allowed to fit behavioural responses but potentially focusses on different aspects of learning. Participants were asked to refrain from liquids for ≥6 h prior to scanning to ensure they were thirsty. The Pavlovian classical conditioning task consisted of four blocks of 30 trials of 8 s each. The task started with one block (30 trials) without juice delivery (the neutral condition), but with the to-be conditioned stimuli (but not yet conditioned). After the neutral block, three blocks followed that included juice delivery. One of two pictures was alternately shown on the screen [the conditioned stimulus (CS)] 2 s after the start of each trial. Two seconds thereafter, the conditioned stimulus was followed by the presence or absence of small amounts (0.2 ml) of rewarding juice [the unconditioned stimulus (US)] at different probabilities (80–20%) (Fig. 1). Every block, a change occurred (three times in total) in which the picture that was ‘rewarding’ (for 80% of the time) was switched with the non-rewarding picture. Before and after the task participants received 0.2 ml fluid after which they were asked how much money they were willing to pay to get more juice (wanting) and how much they enjoyed the taste of the juice (liking). A visual analogue scale ranging from −2 (receive money/unpleasant, respectively) to 2 (pay money/pleasant, respectively) was used to assess wanting and liking, with the centre of the scale being neutral. Juice delivery was via a polythene tube that was attached to a syringe-driver pump (B Braun-Infusomat P) positioned in the scanner control room, interfaced with the stimulus presentation computer. Stimuli were presented using E-prime 2 (Psychology Software Tools, Pittsburgh, PA). The participants were instructed to try to find out which picture predicted the juice delivery and notified that this association could change over time. With changing probabilities of juice delivery, temporal difference reward-learning signals were calculated (Kumar et al., 2008). Other tasks within the same MRI session were carried out after the Pavlovian task to avoid possible confounding effects.

**Data acquisition**

Magnetic resonance images were acquired on a Phillips 3 T Achieva XT MRI scanner using a 32-channel SENSE head coil. T2*-weighted gradient-echo-planar images were collected with the following parameters: repetition time 1500 ms, echo time 28 ms, 25 slices, 1125 volumes, field of view: 240 × 240 mm and matrix 80 × 80; voxel size: 3 × 3 × 3 mm. Slices were oriented with 30° tilt from the AC-PC transverse plane and acquired in ascending order. High resolution T1-weighted anatomical images were acquired with the following parameters: repetition time 8.3 ms, echo time 3.8 ms, 220 slices, field of view: 240 × 188 mm and matrix 240 × 240; voxel size: 1 × 1 × 1 mm. Cardiac and respiratory signals were acquired concurrently during the scan and used to facilitate physiological noise correction in the analysis.

**Data preprocessing**

Images were preprocessed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB R2013a (The MathWorks Inc., Natick, MA). Structural and functional images were reoriented in anterior-posterior commissure alignment to facilitate co-registration. Functional images were realigned to the first functional image and were co-registered to the T1-weighted image. Structural images were segmented into
grey matter, white matter, and CSF. T1-weighted images were used to create a study-specific group template using the DARTEL algorithm (Ashburner, 2007). Subsequently, functional images were normalized to Montreal Neurological Institute (MNI) space using this intermediate group template. Voxel sizes remained $3 \times 3 \times 3$ mm during DARTEL spatial normalization, and images were smoothed with a 4 mm Gaussian kernel. Physiological cardiac and respiratory noise signals were modelled and eliminated retrospectively by the DRIFTER algorithm (Sarkka et al., 2012), a Bayesian method for physiological noise modelling and removal, allowing accurate dynamical tracking of the variations in the cardiac and respiratory frequencies. Frequency trajectories of the physiological signals were estimated by the interacting multiple models filter algorithm (reference signal 1 = respiratory signal: sampling interval = 500 Hz, array of possible frequencies = 10:70 bpm; reference signal 2 = cardiac signal: sampling interval = 500 Hz, array of possible frequencies = 40:140 bpm). The estimated frequency trajectories were then used in a state space model in combination with a Kalman filter and Rauch–Tung–Striebel smoother, which separated the signal into a cleaned activation-related signal, physiological noise, and white measurement noise components. Details regarding this algorithm are described in Sarkka et al. (2012).

**Temporal difference learning model**

From each participant, the E-prime log files were used to extract the timing of the unconditioned stimulus and the conditioned stimulus. All eight time points were modelled, with the conditioned stimulus defined at time point 3 and the unconditioned stimulus at time point 6. The calculation of the temporal difference prediction errors was derived from Kumar et al. (2008), who used a standard temporal difference model derived from Dayan and Abbott (2001). As in previous studies, a same set of parameters was used for all subjects (Kumar et al., 2008, 2018; Daw, 2011; Gradin et al., 2011). The predicted value ($V_t$) at any time $t$ was defined as:

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

where $x_i(t)$ is coded with a 1 or a 0 (for all time points) for the presence or absence of a conditioned stimulus at time $t$. $w_i$ corresponds to a weight that was updated on each trial in order to capture learning by:

$$\Delta w_i = \alpha \sum x_i(t) \delta(t)$$

where $\alpha$ is corresponding to a factor chosen in advance, which represents the learning rate. As recommended for model-based functional MRI analysis (Wilson and Niv, 2015), we selected multiple plausible learning rates from the literature (0.1 and 0.4 from Kumar et al., 2008 and O’Doherty et al., 2006; 0.2 from O’Doherty et al., 2003, 2004; 0.45 from Gradin et al., 2011; 0.5 from Lawson et al., 2017) and explored which learning rate fitted our data best. We chose $\alpha = 0.45$ as the optimal learning rate based on optimal signal-to-noise ratio calculations and estimation of efficiency values of SPM designs (Liu et al., 2017 and Supplementary material for details regarding the calculation of estimation efficiency). To ensure
our results were robust, we compared temporal difference (TD)-related activation in the CS × TD + US × TD contrast across the range of learning rates (Supplementary material).

The temporal difference error signal was defined as:
\[ \delta(t) = r(t) + \gamma \hat{V}(t + 1) - \hat{V}(t) \]
where \( r(t) \) is coded with a 1 or a 0 (for all time points) for delivery of juice or no-juice, respectively and \( \gamma \) corresponds to a factor chosen in advance, which determined the importance of later reinforcements compared with previous ones. Following previous studies, \( \gamma = 1.0 \) was used (Kumar et al., 2008; Gradin et al., 2011). This means that the model did not include discounting effects and assumed that such effects did not differ between groups, which is a common assumption in model-based functional MRI literature (O’Doherty et al., 2003, 2006; Kumar et al., 2008; Gradin et al., 2011).

### Statistical analysis

#### Sample characteristics

Analyses were performed with SPSS v22.0 (SPSS Inc., USA). We used \( P < 0.05 \) as threshold for significance. Independent sample \( t \)-tests, \( \chi^2 \)-tests and non-parametric Mann-Whitney U-tests were used to compare demographics (age, sex, education, IQ) and clinical variables (HDRS, SHAPS, number of lifetime episodes, age of onset) between rrMDD and healthy control subjects.

#### Behavioural data

Group differences in wanting and liking ratings were analysed using repeated-measures analysis of variance with group (rrMDD, healthy controls) as the between-subjects factor and time (pre-task and post-task) as the within-subjects factor. Because groups differed slightly but significantly, we used HDRS scores as a covariate, to exclude effects driven by (small) HDRS differences.

#### Imaging data

In SPM12, an event-related random effects design was used for the analysis. For each participant, first-level haemodynamic responses for each stimulus (conditioned and unconditioned) were modelled using a canonical haemodynamic response function model. The temporal difference prediction errors were entered into the model as parametric modulators for the conditioned and unconditioned stimulus conditions. To look at main cue and delivery task effects separately, we modelled a conditioned stimulus > neutral and a unconditioned stimulus > neutral condition. We also modelled a pooled contrast (conditioned stimulus + unconditioned stimulus > neutral) to see if the task would elicit ventral striatum activity regardless if it was during cue (conditioned stimulus, CS) or delivery (unconditioned stimulus, US). Given our primary hypothesis about temporal difference (TD) related activation, we modelled the contrast CS × TD + US × TD. Separate contributions of the conditioned and unconditioned stimulus temporal difference errors were also modelled by a CS × TD and US × TD condition. A high-pass filter of 128 s was used to remove low frequency noise. Realignment parameters and their first derivatives were added to the model to address residual movement not corrected by realignment.

### Results

#### Patient disposition and sample characteristics

From the 62 rrMDD patients and 41 healthy control subjects that were scanned, we excluded three patients and two healthy controls because of abnormal brain anatomy and five patients and four healthy controls because of corrupted or missing task data. During the analysis phase, 18 patients and eight healthy controls were excluded because of missing or corrupted physiological data needed for filtering of cardiac and respiratory noise, leaving a sample of 36 patients and 27 healthy controls included in the final analyses. Excluded subjects did not significantly differ in sample characteristics from the included sample. No significant differences were observed between rrMDD patients and healthy controls (Table 1), except higher residual symptomatology (HDRS; \( U = 224, P < 0.001 \)) and anhedonia (SHAPS; \( U = 253, P = 0.002 \)) in rrMDD patients.
Behavioural results

For the wanting and liking ratings (corrected for HDRS differences) no main effect of group or time was observed. No significant group x time interactions were identified (Fig. 2).

Functional MRI results

We observed main effect activation of the ventral striatum during delivery of cues and reward (conditioned stimulus + unconditioned stimulus > neutral, conditioned stimulus > neutral and unconditioned stimulus > neutral contrasts) (Table 2 and Supplementary Fig. 2). We also found a main effect of prediction error in the ventral tegmental area and the ventral striatum (CS/C2 TD + US/C2 TD contrast) (Table 2 and Supplementary Fig. 3). We found increased temporal difference-related activation (CS x TD + US x TD contrast) in the ventral tegmental area in rrMDD patients compared to healthy controls ($P_{FWE,SVC} = 0.028$) (Table 3 and Fig. 3). The significance of this group difference was $P_{FWE,SVC} = 0.048$ after correction for HDRS scores between groups (Supplementary Fig. 4). Temporal difference signals in the ventral striatum did not differ significantly between groups.

When comparing rrMDD versus healthy controls in the CS/C2 TD and the US/C2 TD contrast separately, differences in temporal difference-related ventral tegmental area activation were not significant (Table 3).

Association between ventral tegmental area temporal difference signal and anhedonia ratings

The regression model with SHAPS scores, group, group x SHAPS interaction and HDRS explained 21% of the variance [$F(4,57) = 3.78$, $P = 0.009$]. This
model showed a significant group × SHAPS interaction \([t(57) = -2.29, P = 0.026]\) in addition to the main effect for group \([t(57) = 3.03, P = 0.004]\) (Fig. 4). In rrMDD patients, higher anhedonia was associated with lower ventral tegmental area temporal difference activation. In healthy controls, higher anhedonia was associated with higher ventral tegmental area temporal difference activation.
Discussion

This study explored the response of the ventral tegmental area and ventral striatum during a classical conditioning functional MRI task in medication-free patients with rrMDD compared to healthy control subjects. We found significantly increased temporal difference reward learning activation in the ventral tegmental area in rrMDD patients compared to healthy controls. No differences between the groups were observed for ventral striatum activity. Moreover, we investigated the relationship with anhedonia and showed that in rrMDD patients, higher anhedonia was associated with lower ventral tegmental area temporal difference reward learning activation, while in healthy controls, higher anhedonia was associated with higher ventral tegmental area activation.

This study did not demonstrate the difference in basic wanting and liking processing, as described in depressed patients (Treadway and Zald, 2011). Furthermore, wanting and liking properties did not differ over time between both groups. This result is in agreement with McCabe et al. (2009), who also found no significant differences between recovered depression patients and healthy controls on ratings of wanting (pleasantness) and liking. This suggests that these differences are either not present, or are smaller in a remitted state. This notion is further corroborated by our functional MRI findings, where we found no group differences in basic processing of reward in the ventral striatum. Previous functional MRI studies in depressed patients found reduced ventral striatum activity (Pizzagalli et al., 2009; Smoski et al., 2009; Robinson et al., 2012), although not consistently (Knutson et al., 2008; Rothkirch et al., 2017; Rutledge et al., 2017). Inconsistencies might be attributable to differences in study designs and/or patient characteristics. However, studies investigating reward processing in remitted depression patients, consistently never reported ventral striatum differences (Dichter et al., 2012; Ubl et al., 2013; Hammar et al., 2016). We therefore propose that the reduction in reward sensitivity and ventral striatum activation during reward delivery in depressed patients is likely to recover after achieving remission and therefore could be considered a state effect. Another explanation for a difference between ventral tegmental area and ventral striatum temporal difference activation can be based on findings by Klein-Flügge et al. (2011), who demonstrated that classic temporal difference reward prediction error activity was specific to the ventral tegmental area, but not the ventral striatum, which suggests decoupling between ventral tegmental area dopaminergic neuron firing and ventral striatum dopamine release.

In contrast to the suggested recovery of basic wanting and liking processing in patients with remitted depression, our results show that the underlying learning signals to learn the associations between reward outcome and stimuli are impaired. Kumar et al. (2008) demonstrated increased ventral tegmental area temporal difference-related activations during reward-learning in patients while depressed, which correlated with illness severity. These findings were interpreted as reflecting a compensatory response to an impaired function of other non-brainstem regions, such as the ventral striatum, of the mesolimbic pathway. However, the current results demonstrate that also in remitted recurrent depression, increased ventral tegmental area activity during reward-learning persists, while the difference in temporal difference-related activation in the ventral striatum seems to be restored.

However, Kumar et al. (2008) investigated a sample of depressed patients who were non-responsive to long-term antidepressants, and healthy control subjects in unmedicated and (acutely) medicated state. Interestingly, the temporal difference signals in the ventral striatum of medicated healthy controls (compared to the unmedicated healthy controls) were reduced and did no longer differ significantly from patients with MDD. Animal studies report different effects of acute versus chronic administration of antidepressants (Sekine et al., 2007) and in patients with MDD, acute administration of antidepressants reduced temporal difference error-related neural activity in the ventral striatum (McCabe et al., 2010; Chase et al., 2013; Herzallah et al., 2013). Therefore, it could be hypothesized that reduced temporal difference signals in the ventral striatum in medicated, depressed patients might reflect medication effects instead of state effects. Indeed, a recent paper corroboratively reported no differences in prediction error-related activity in the ventral striatum in unmedicated depressed patients versus healthy control subjects (Rothkirch et al., 2017). We are aware that there are relatively few studies on unmedicated samples, and that previous cohorts are often slightly less severe than medicated cohorts. Therefore, it is difficult to make claims about medication based on the present unmedicated cohort, and more direct

Figure 4 Association of ventral tegmental area activation and anhedonia (SHAPS). Significant group × SHAPS interaction [t(57) = −2.29, P = 0.026] and a main effect for group [t(57) = 3.03, P = 0.004]. HC = healthy controls.
comparisons are needed. However, the described effects of medication could provide an additional explanation for our findings of comparable temporal difference-related activity in the ventral striatum.

Our finding of increased ventral tegmental area temporal difference signals in rrMDD patients versus healthy control subjects is in line with the report in unresponsive medicated patients with MDD (Kumar et al., 2008) and suggests a trait-like abnormality, i.e. impaired reward-related learning is associated with MDD, and seems to be state-independent, which are both important criteria of the endophenotype concept (Gottesman and Gould, 2003), relevant for recurrent depression. Nevertheless, to the best of our knowledge, the heritability (another endophenotype characteristic) of impaired reward-related learning has yet to be demonstrated.

The phasic dopamine firing into temporal difference signals has been well described (Schultz et al., 1997; Schultz, 1998; Tobler et al., 2005), which makes it valid to interpret temporal difference signal impairments as a dysfunction of the dopaminergic system. The role of the (dysfunctional) dopamine system in the pathophysiology of MDD has been emphasized by Dunlop and Nemeroff (2007). They suggest the existence of subtypes of depression stemming from abnormal dopaminergic neurotransmission, and suggest further research regarding the involvement of dopaminergic circuit dysfunction in non-response to treatment, or treatment resistance. Given that 20% of recurrent depressive episodes become chronic despite treatment (Judd et al., 1998), and with the present findings in mind, future studies focusing on reward-related learning impairments in treatment-resistant depression are warranted.

The significant group x anhedonia interaction indicated that rrMDD patients with higher levels of anhedonia have reduced ventral tegmental area temporal difference signals. Reduced ventral tegmental area activity was also reported by Dillon et al. (2014a), who investigated reward memory in unmedicated adults with MDD. Furthermore, the group x anhedonia interaction indicated that healthy controls with higher levels of anhedonia have increased ventral tegmental area temporal difference signals. Interestingly, a study in healthy participants reported that higher levels of anhedonia were not associated with the ventral tegmental area, but instead associated with reduced activity in other key areas of the reward circuitry linked to the ventral tegmental area (basal forebrain, ventral striatum). Therefore, the observed increased ventral tegmental area activity in healthy controls might be compensatory to overcome a diminished reward sensitivity in more anhedonic healthy controls (Keller et al., 2013).

In contrast, the opposite relation between anhedonia and ventral tegmental area temporal difference activation in MDD, even in the remitted state, could be interpreted in accordance with Eldar and Niv (2015), who suggested that reward prediction errors are strongly related to mood. If remitted depressed individuals are recovering from depression, it may be that they experience larger positive prediction errors as they find rewarding events more rewarding than they are used to. Hence a larger reward prediction error might be observed. This would explain why remitted depression patients with greater residual anhedonia have smaller prediction error responses.

Another explanation can be based on Liu et al. (2017), who found that in depressed, unmedicated MDD, especially in response to expected punishment, higher levels of anhedonia were associated with attenuated habenula activation. The habenula is not only important in punishment processes (i.e. expectation of aversive stimuli), but also plays a central role in reward processing (i.e. absence of rewards) (Lawson et al., 2014), specifically via projections to the ventral tegmental area. Studies investigating habenula function in humans and animal models of MDD showed that the habenula is hyperactive in MDD (Shumake and Gonzalez-Lima, 2013; Dillon et al., 2014b; Lecca et al., 2014; Benarroch, 2015; Zhao et al., 2015; Liu et al., 2017). As the habenula is known to inhibit ventral tegmental area dopaminergic firing (Matsumoto and Hikosaka, 2007), and the absence of a reward is in particular a strong activator of the habenula (Proulx et al., 2014), this could explain the negative correlation between anhedonia and ventral tegmental area temporal difference signals in rrMDD patients. More anhedonic rrMDD patients, experiencing less/absence of rewards, might have further increased habenula hyperactivity, resulting in increased (habenula-driven) inhibition of dopaminergic firing in the ventral tegmental area. By a stronger decrease in reward expectancy this could even strengthen anhedonia and associated depressive behaviour in a vicious cycle. Via this mechanism, anhedonia might have a modifying effect on the effectiveness of behavioural treatments, commonly used to alleviate MDD, which, however, remains to be established (Treadway and Zald, 2011). Notably, in rats, a decrease of habenula firing has been associated with reduction of depressive-like behaviour (Li et al., 2011), and deep brain stimulation in the habenula resulted in remission of symptoms in a patient with treatment-resistant depression (Sartorius et al., 2010). Unfortunately, due to low power, our present study design was not suitable to specifically explore negative temporal difference errors coding for the absence of a reward. Therefore, the role of the habenula in the association between anhedonia and temporal difference signals remains speculative, requiring verification in future studies.

Regardless whether a functional impairment of the ventral tegmental area or the habenula underlies the association with anhedonia, it would be interesting to investigate whether the observed impairments in reinforcement learning are associated with recurrence. A link between recurrence and impaired reinforcement learning would suggest that—in line with previous research—the focus of therapy should not only lie on diminishing negative affect but also enhancing positive affect by training patients to focus attention on positive reinforcers (Wichers et al., 2010, 2012; Servaas et al., 2017).
Focusing on positive experiences might train the ability to make associations between behaviour and pleasurable outcomes and might reinforce repetition of reward-provoking behaviour (operant conditioned learning). Training the ability for (rr)MDD patients to learn about rewarding feedback in daily life and remediate impaired reinforcement learning should be investigated in future studies, while considering anhedonia as a moderator.

Strengths and limitations

This is the first study exploring reinforcement learning during remission in a relatively large group of unmedicated patients with MDD. Nevertheless, potential limitations are present. First, as in the original task (Kumar et al., 2008), the experimental task lacked an active response to the appearance of the pictures on the screen. This excludes the possibility of any behavioural confound in the Pavlovian learning. Although this passive conditioning task was specifically used to assess particular aspects of learning, participants might have lost their engagement or attention to the task and we were not able to assess individualized learning rates. In new experiments, an active response (e.g. button press) will be embedded in the task, which will facilitate the possibility to fit the model to the data and select parameters that show the best overall fit to the signals. Furthermore, future analyses could benefit from novel methods that extract parameters by fitting computational models to neural data alone or to a combination of behavioural and neural data at the same time (Purcell et al., 2010; Turner et al., 2013; Frank et al., 2015; Turner et al., 2016; van Ravenzwaaij et al., 2017). Second, the direct measurement of dopamine signalling with functional MRI is impossible. Nevertheless, strong evidence supports that blood oxygen level-dependent signals in reward-related brain areas reflect dopaminergic release (Pessiglione et al., 2006; Knutson and Gibbs, 2007). Third, by modelling the temporal difference error signal and comparing patients and controls, we reject the null hypothesis of no differences between groups. These differences between groups could be due to either actual difference in dopaminergic learning signals between groups, or differences between groups (and individuals in the groups) in learning rate and/or discount factor, which are used to model the temporal difference errors. However, previous research found no differences in model parameters between patients with MDD and healthy controls (Gradin et al., 2011). Moreover, using a single set of model parameters across all participants and groups showed more robust results in multi-subject functional MRI studies (Daw, 2011). Therefore, we interpret our findings as representing differences in dopaminergic temporal difference signals between groups. A fourth limitation is that the a priori choices that were made for our analysis (e.g. learning rate selection, choice of smoothing kernel) are one out of many approaches that can be considered. We chose to explore plausible learning rates from literature instead of exploring an entire range of learning rates between 0 and 1. This method was chosen because the primary aim was to investigate the difference between patients and controls and not to methodologically explore how to model learning rates. Furthermore, it has been suggested in literature that even gross deviations in the learning rate lead to only minimal changes in the neural results and that precise model fitting is not always necessary for model-based functional MRI (Wilson and Niv, 2015). When exploring our neural results in the range we described, we indeed found comparable results when using different learning rates. A fifth limitation is that a currently depressed group or scanning of the subjects when depressed was not incorporated in the present analysis. This hampers the ability to draw inferences about persistence. However, in its present form, the study can be very helpful for the identification of factors that remain impaired during remission in depressive patients with a history of recurrence. Lastly, no individual levels of thirst were obtained at the start of the experiment. Nevertheless, participants confirmed that they refrained from liquids for \( \geq 6 \) h prior to scanning, which made it fair to assume sufficient levels of thirstiness.

Conclusion

In summary, we demonstrated impaired reward-related learning in unmedicated patients with a recurrent MDD during remission, which may be an (endo)phenotype linked to depression vulnerability. Our findings add to evidence for state-independent, impaired temporal difference learning signals in the ventral tegmental area, which requires further investigation as an endophenotype for (recurrent) MDD. Furthermore, the association between impaired reinforcement learning and anhedonia in rrMDD patients strengthens the need to focus on this residual symptom and investigate remediation of hedonic capacity and processing of reward-related learning in rrMDD.

Acknowledgements

We thank three anonymous reviewers for their thoughtful comments which clarified our methods and suggested us to perform additional sensitivity analyses.

Funding

This work was supported by unrestricted personal grants from the AMC to R.J.T.M. (AMC PhD Scholarship) and C.A.F. (AMC MD-PhD Scholarship), and a dedicated grant from the Dutch Brain Foundation (Hersenstichting Nederland: 2009(2)-72). H.G.R. is supported by an NWO/ZonMW VENI-Grant #016.126.059. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data;
preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at Brain online.

References


