Avoiding monetary loss: a human habenula functional MRI ultra-high field study.

Kathrin Weidacker [1,2], PhD, Seung-Goo Kim[1,3], PhD, Camilla L. Nord[1,4], PhD, Catarina Rua[5], PhD, Christopher T. Rodgers[5], DPhil, Valerie Voon[1], MD PhD

[1] Department of Psychiatry, University of Cambridge, Box 189, Level E4, Addenbrooke's Hospital, Cambridge, CB2 0QQ, United Kingdom.

[2] Department of Psychology, Swansea University, Singleton Campus, Swansea, SA2 8PP, United Kingdom.

[3] Department of Psychology and Neuroscience, Duke University, Durham, North Carolina, United States

[4] MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, CB2 7EF, United Kingdom

[5] Wolfson Brain Imaging Centre, University of Cambridge, Box 65, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom.

Corresponding author: **Valerie Voon**, University Department of Psychiatry, Box 189, Level E4, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; Email: vv247@cam.ac.uk, telephone: +44 1223 761327, fax: +44 1223 336968

Abstract

A number of convergent human neuroimaging and animal studies suggest that habenula neurons fire in anticipation of non-rewarding outcomes, and suppress their firing in anticipation of rewarding outcomes. This normative function of the habenula appears disrupted in depression, and may be critical to the anti-depressant effects of ketamine. However, studying habenula functionality in humans using standard 3T MRI is inherently limited by its small size. We employed ultra-high field (7T) fMRI to investigate habenular activity in eighteen healthy volunteers during a Monetary Incentive Delay Task, focussing on loss avoidance, monetary loss and neutral events. We assessed neural activation in the field of view (FOV) in addition to ROI-based habenula-specific activity and generalized task-dependent functional connectivity. Whole FOV results indicated substantial neural differences between monetary loss and neutral outcomes, as well as between loss avoidance and neutral outcomes. Habenula-specific analyses bilateral deactivation during loss avoidance, compared to other outcomes. This first investigation into the habenula's role during loss avoidance revealed that the left habenula further differentiated between loss avoidance and monetary loss. Functional connectivity between the right habenula and the ipsilateral hippocampus and subcallosal cingulate (regions implicated in memory and depression pathophysiology) was enhanced when anticipating potential losses compared to anticipating neutral outcomes. Our findings suggest that the human habenula responds most strongly to outcomes of loss avoidance when compared to neutral and monetary losses, suggesting a role for the habenula in both reward and aversive processing. This has critical relevance to understanding the pathophysiology of habenula function in mood and other neuropsychiatric disorders, as well as the mechanism of action of habenula-targeting antidepressants such as ketamine.

Keywords: habenula, reward processing, loss avoidance, monetary incentive delay task, functional connectivity

1. Introduction

The human habenula (HB) complex plays a critical role in anti-reward processing and is relevant also, to reward processing [1, 2]. The lateral habenula (LHB) has been implicated in the pathophysiology of psychiatric disorders such as Major Depressive Disorder (MDD)[3-6], with rodent studies highlighting the LHB in mediating the anti-depressive effect of ketamine[7]. The human HB is located bilaterally between the pineal gland and the dorsomedial thalamus and represents a small neural structure (around 30 mm^3 volume per hemisphere)[8] composed of grey and white matter. Cytologically, the HB can be subdivided into medial and lateral nuclei[9] which have different genetic profiles as well as anatomic connections based on non-human studies[10]. The capacity to differentiate medial and lateral HB in human *in vivo* neuroimaging research is limited due to their small size[11]. Here we aim to assess HB function using a task-based ultra-high field 7T fMRI study in healthy volunteers focusing on loss avoidance and incurring losses.

Non-human primate research suggests the LHB plays a critical role in the upstream modulation of midbrain dopaminergic neurones and is involved in anti-reward processing and, to a lesser extent, in reward processing[1, 2]. When non-human primates are anticipating non-rewarding outcomes, excitation of the LHB neurons temporally precedes the inhibition of dopamine neurons. Similarly, electric stimulation of LHB neurons induced inhibition of dopamine neurons, indicating that the inhibitory effect of non-rewarding stimuli on dopamine neurons is guided by the LHB. The anticipation of rewarding outcomes, on the other hand, leads to decreased firing of LHB neurons[2]. A similar neuronal response profile is seen in response to the outcomes themselves, with strong excitation of LHB neurons for negative and inhibition for positive outcomes[1]. Single-cell recordings further showed that the neuronal response of the LHB depends on the context, with strongest neuronal responses to cues predicting the worst outcome among the available alternatives (for instance: the absence of reward when the alternative is reward, or the presence of punishment when the alternative is absence of punishment)[1]. The LHB neurons are also sensitive to the mismatch between prediction and outcome as indicated by weaker excitatory responses when a

negative outcome was fully predictable compared to a less certain outcome, and, greater negative prediction errors being associated with increased LHB neuron firing rate[1]. Thus, the LHB is responsive to both reward and negative prediction errors, and in particularly to negative motivational value, with the encoding direction opposite to that of dopamine neurons[1, 2].

In humans, neuroimaging research using task-based functional magnetic resonance imaging (fMRI) to delineate the HB function is sparse due to the inherent difficulty of isolating signal change from small structures such as the habenula. Similarly, the majority of fMRI studies on the human HB were carried out using standard field strengths, such as 3T, which further limits the delineation of the HB due to lower resolution than recent advantages utilizing ultra-high field imaging (7T). Indeed, the standard voxel size used in 3T EPI sequences for fMRI studies is 3 mm, a major limitation for a nucleus of 30 mm^3 volume. Despite these constraints, previous human fMRI studies (including those using specialised high-resolution 3T sequences) converge with primate single-cell recordings demonstrating the importance of the human HB for the anticipation of punishment and changing reward contingencies.

In healthy humans, anticipating electrical shocks (versus neutral outcomes) evoked increased activation in bilateral insula, caudate, but also in the left, and to a lesser extent, in the right, HB[12]. Region of interest (ROI) analyses revealed that the left-hemispheric HB activity increase during punishment anticipation also holds when compared to reward (juice receipt) anticipation; anticipating rewarding versus neutral outcomes did not affect left HB signal change[12]. In line with primate findings, the human HB is also sensitive to probabilities. For example, cues predicting a high versus low chance of losing points in a guessing game evoked increased left-, but not right-, hemispheric HB activation[13]. While this previous study found the left HB to be sensitive to probabilities, a separate study investigating HB responses to cues indicative of a high versus low chance for upcoming punishment (in the form of electrical shocks) or monetary rewards reported a bilateral increase in HB activation for the anticipation of punishment and a decrease for monetary rewards. Interestingly, this study also investigated monetary losses within the same design and while

HB activation increased, the HB response to losses fell in between monetary wins and shock as punishment, in line with earlier reports of the HB responding to the most salient among outcomes[1, 14].

Converging with the HB's role in tracking prediction errors in non-human primates, human *in vivo* neuroimaging research suggests that right HB activation increases linearly in response to increased adversity of anticipatory cues, highlighting the sensitivity of the human HB to the motivational value of anticipatory cues[14]. When utilizing losing and winning points as punishment and reward conditions, no laterality effect was observed regarding prediction errors, instead, bilateral HB activation was enhanced for punishment-related prediction errors compared to reward-based prediction errors[15]. Investigating neural responses to prediction errors across studies, a recent Activation Likelihood Estimation (ALE) meta-analysis confirmed a role for the human HB particularly for punishment-related prediction errors, in addition to brain areas such as the middle frontal gyrus (MFG) and the insula. Reward-based prediction errors on the other hand, were associated with activation changes in reward-processing related brain areas such as the striatum including the nucleus accumbens (NACC)[16].

While the reviewed findings relate to the anticipation of positive and negative outcomes, differential HB activation has also been observed during outcome presentation. When assessing HB responses to outcomes in a guessing game, bilateral HB responses increased for monetary losses over wins, while only left-hemispheric signal change differed between healthy volunteers and patients suffering from MDD[13]. In contrast, when investigating HB response to outcomes such as loosing or winning points (each normalized to neutral outcomes), left-hemispheric HB activation differentiated between punishment, which led to increased, and reward, which resulted in decreased HB activation[15]. Regarding the presentation of negative and positive non-verbal feedback during a prediction task, enhanced bilateral thalamus activation, which included the HB, was reported for negative versus positive feedback alongside greater activity in bilateral anterior insula and anterior cingulate cortex (ACC)[17].

In summary, convergent studies of human HB function using 3T highlight HB responsiveness to punishment and reward, although laterality effects remain unclear. During anticipation, as well as during outcome presentation, the human HB increases activation for stimuli associated with punishment over reward. This is convergent with non-human primate research investigating the firing patterns of LHB neuron populations, which showed that HB subpopulations are activated following negative and inhibited following positive motivational stimuli[1].

While task designs differed across studies, most investigations utilized some form of delayed anti-/reward paradigm, in which negative outcomes are related to losing points, money or punishment. However, whether the human HB also responds differentially to loss avoidance, the absence of incurring losses, has not yet been investigated. Neurally, the anticipation of loss avoidance differs from that of reward anticipation, with anticipation of loss avoidance evoking less signal change in the NACC in children[18]. Similarly, while ventral striatal activation was generally decreased for the anticipation of loss avoidance compared to rewards in healthy adolescents, adolescents with Attention-deficit/Hyperactivity disorder (ADHD) expressed reduced ventral striatal activation during reward anticipation, but not during loss avoidance anticipation when compared to controls, further highlighting the differences underlying reward and loss avoidance during anticipation[19]. Additional differences between loss avoidance and reward were reported during the outcome phase. In adults, neural responses to feedback relating to the successful avoidance of losses versus neutral outcomes increased activation in the inferior frontal gyrus and the cerebellum while clusters being more activated to rewarding than neutral outcomes were more widespread, including caudate, globus pallidus, and cingulate brain regions among others[20]. Similarly, neural activity during loss avoidance, but not reward outcomes, differentiated between controls and adults unmedicated for childhood ADHD, with reduced bilateral insular and precentral gyrus activity in the latter group for loss avoidance outcomes[21].

Previous research comparing successful loss avoidance to loss incurrence, instead of reward, did not focus on HB activation, but uncovered enhanced activation for loss avoidance in superior

temporal gyri, pre/-cuneus, and reward-related subcortical brain regions such as the caudate and the ventral striatum in healthy adults[22]. The concept of loss avoidance is further of relevance for disorders of addiction. Comparing loss avoidance to loss incurrence between alcohol-dependent patients and controls revealed enhanced activation in reward-related areas such as ventral striatum, caudate, and putamen as well as in insula, temporal gyri, MFG and precuneus in controls, while neural activation during reward trials did not differ between groups[22]. A separate investigation confirmed the reduced striatal activation in alcohol-disordered patients during successful versus non-successful loss avoidance, and additionally revealed aberrant loss avoidance processing in pathological gamblers, with reduced activation in ventral striatum and medial prefrontal cortex compared to controls[23].

Loss avoidance is of high relevance to mental disorders such as ADHD, but also plays and important role in addiction, particularly for alcohol abuse and gambling disorders. Neurally, loss avoidance differs from reward processing during anticipation and receipt, and to incurring losses during the feedback phase. Experimentally, most reviewed research utilized versions of the Monetary Incentive Delay (MID) paradigm[24]. During each trial, a cue is presented which indicates the upcoming trial type, e.g. punishment or reward. This is followed by a variable delay during which anticipation of the trial outcome occurs. Thereafter participants are asked to correctly perform a task, for instance, to press a button corresponding to the direction of an arrow. This response phase is typically dynamically adjusted to enhance task difficulty and to increase the proportion of errors (e.g., [25-28]). Thereafter, visual feedback relating to the correctness of their response is presented, termed outcome phase. The MID is especially suited for studying the neural basis of anticipation and receipt due to separating the anticipation and outcome phases in time.

Given the importance of loss avoidance for psychiatric disorders, we here report the first results on human HB function in relation to loss avoidance in healthy volunteers, making use of advanced neuroimaging techniques, ultra-high field imaging (7T). We utilize the MID task with Loss and Neutral trials, separating monetary loss avoidance and loss incurrence during the outcome

phase. We additionally investigate generalized task-dependent functional connectivity involving the HB.

2. Methods

2.1 Participants

Twenty right-handed participants (10 male) participated in the study. All participants fulfilled 7T scanning safety criteria and reported no previous brain injuries, seizures, and mental health diagnoses. All participants provided informed consent before participation and the study was approved by the local ethics committee. Two females were excluded, one due to the HB moving out of the field of view (FOV), the second due to technical issues. The resultant sample of eighteen participants had a mean age of 29.78 years, age ranged between 20 and 42 years. Sample size was determined based on previous 7T studies and inclusion/exclusion criteria were established prior to data analyses.

2.2 Monetary Incentive Delay (MID) Task

The MID task was programmed using Presentation software (version 20.2, www.neurobs.com) and responses were recorded using an fMRI-compatible button box. The MID task consisted of two trial types: 40 Loss and 30 Neutral trials, shown in equal proportions in the first and second half of the task. Trial presentation was pseudo-randomised with the restriction of less than three subsequent trial type repetitions. Each trial consisted of fives phases: cue, anticipation, response, blank screen, and outcome phase (see Figure 1 for stimuli and durations). Participants were instructed that incorrectly responding during Loss trials results in losing money, and that if they do not lose much money during this task, they will receive an additional £5 at the end of the experiment, in addition to their payment for participation which was £10 per hour. For Neutral trials, participants were instructed that correctness of response does not affect their pay. In reality, all participants received the additional £5 at the end of the experiment. In the first part of the trial, the

Cue phase, participants were shown a visual cue (red square with a crossed out £ sign for Loss trials; Yellow triangle with a dot in the centre for Neutral trials) indicating the upcoming trial type. During the anticipation phase, a slightly altered version of the Cue was shown (empty red square for Loss trials; Empty yellow triangle for Neutral trials). During the response phase, an arrow was either pointing to the left (requiring a button press with the right index finger) or to the right (requiring a button press with the right middle finger). Left and right arrows were presented equally often per trial type, pseudo-randomised with the restriction of less than four trials requiring the same response in a row. All responses were required to be as fast and accurate as possible. To enable analyses of correct (pressing the correct button within the response time window) and wrong responses (pressing the wrong button within the response time window, misses, or any responses outside the response window), the duration of the response window was constantly adjusted to yield 50 % correct responses, in line with previous MID task designs[25-28]. The initial response window was set to 250 ms and adjusted using an independent staircase procedure per trial type. Incorrect responses prolonged the response window by 50 ms, correct responses decreased the allowed response duration by 50 ms. Due to the staircase procedure, button presses occurring outside the response window were likely. To avoid contamination of the outcome phase by button presses, a blank screen (500 ms duration) interspersed the response phase and the outcome phase. What was shown during the outcome phase of Loss trials depended on the correctness of the response during the response phase: Correct button presses led to showing “Money not affected”, indicating successful loss avoidance, and wrong/late responses led to a screen with a crossed-out pound coin accompanied by “You lost money”, indicating monetary loss. The outcome phase for Neutral trials was independent of the correctness of responses and led for both correct and wrong/late button presses to the “Money not affected” screen.

Between trials, a fixation cross was shown. The duration of the fixation cross, as well as the duration of the anticipation phase, were each drawn randomly from two independent (one for fixations and one for anticipation) discrete uniform distributions (each from 2000 to 5000 ms, in

steps of 50 ms) without replacement, except for 9 additional symmetrical samplings (the mean: 3500 ms, the four shortest: 2000, 2050, 2100, 2150 ms, and four longest durations: 4850, 4900, 4950, 5000 ms) to conform with the number of trials.

2.3 Image Acquisition

Scanning was performed using the 7T Terra MRI scanner (Siemens, Erlangen, Germany) at the Wolfson Brain Imaging Centre, Cambridge, UK and a 32-channel receive (1Tx/32Rx) head coil (Nova Medical Inc, MA, USA).

To obtain a high-quality uniform T1w image, the Magnetization Prepared with 2 Rapid Gradient Echoes (MP2RAGE) sequence was used[29] with the following parameters: TR/TE = 4300/1.99 ms, TI1/TI2 = 840/2370 ms, nominal FA1/FA2 = 5/6°, in-plane resolution = 0.75 x 0.75 mm^2, 0.75 mm slice thickness, image matrix = 300 x 320, 224 slices, GRAPPA acceleration factor = 3, bandwidth = 250 Hz/pixel.

The functional data was acquired using a 0.8 mm isotropic 2D single-band gradient-echo echo-planar imaging (GE-EPI) sequence: TR/TE = 3000/22 ms; nominal FA = 77°, 36 slices, no slice gap, image matrix = 256 x 256, GRAPPA acceleration factor = 3, bandwidth = 1028 Hz/pixel, phase-encoding direction anterior-posterior (A-P), partial Fourier = 5/8. Five volumes were collected with the same parameters as the functional scan but with phase encoding reversed (P-A) before task onset for B_0 distortion correction. During scanning, the MP2RAGE T1 maps were visualised for each individual to localise the HB. The FOV was angled for each subject to ensure the HB was located at least 5 slices above the lower border of the FOV and that the FOV extended as far as possible into the ventromedial prefrontal cortex (see Figure 2), as such the FOV tilt angles varied minimally across participants.

To correct for physiological fluctuations in the fMRI data, cardiac and respiratory data were recorded from the scanner's pulse oximeter on the left index finger and a pneumatic belt around the diaphragm.

2.4 Image Pre-processing

SPM12 (Welcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm) was utilized for all preprocessing steps unless specified otherwise. Structural images were skull-stripped and bias-corrected using SPM12's unified segmentation approach[30] and normalised to MNI space. For creation of the study-specific 7T template, all normalised T1w images were averaged using SPM12's 'imcalc'.

For the fMRI data, field maps were created using FSL's 'topup' routine (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TOPUP) to reduce geometric distortions based on five images of each phase encoding direction. The functional data were realigned and unwarped, slice timing corrected, the mean images were bias-corrected using the unified segmentation approach[30] and coregistered to the individual's T1w in native space.

For the HB ROI analyses, separate left and right HB masks were created by manually selecting each individual habenular voxel on each individual's T1w volume (in native space) using the software MRIcron[31]. Given the high contrast of the 7T MP2RAGE, HB delineation was based on visual inspection of T1w image intensity. We additionally created HB ROI's based on the geometric method[8], outlined in the supplement and Supplementary Figure 1, as additional support for our results. A visual example of the derived masks based on image intensity and the geometric method is presented in Supplementary Figure 2 in addition to the supplementary results based on the geometric method. The percentage of HB voxels falling outside the participants' task fMRI FOV was minimal, ranging between 0 and 7.38% (M = 1.06, SD = 2.39) for bilateral HB masks based on image intensity and between 0 and 7.56% (M = 1.08, SD = 2.25) for bilateral HB masks created via the geometric method.

Following HB delineation, the fMRI data subjected to ROI analyses were smoothed using a 2 mm FWHM, whereas the fMRI data used for analysing the whole FOV were normalised to MNI space

and subsequently smoothed with a 6 mm FWHM. The usage of different smoothing kernels is in line with previous research on HB activity[8, 14].

Additional motion regressors were created using the Artefact Detection Toolbox (ART, http://www.nitrc.org/) with cut-offs reflecting the 97th percentile which are suited for our voxel size (global signal change > 5, translation > .9 mm). For the creation of physiological regressors relating to cardiac and respiratory effects, TAPAS R2019b as implemented within the MATLAB PhysIO Toolbox[32] was utilized. The first level analyses additionally incorporated the six motion regressors and the changes in translation and rotation between subsequent volumes. The first level model included each MID component (Cue Loss, Cue Neutral, Blank, Reminder Loss, Reminder Neutral, Arrow, Outcome Loss Avoidance, Outcome Loss, Outcome Neutral, Fixation) and all events were modelled with a boxcar function.

2.5 Statistical Analyses

The whole FOV fMRI analyses were carried out per experimental MID phase (Cue, Anticipation, Outcome) and based on the activation differences (e.g. Anticipation phase: Loss - Neutral) calculated at the first level. These contrasts were then included at the group level to perform one-sample *t*-tests, which were thresholded at $p_{uncorrected} < .001$. Statistical differences were defined as $q < .05$ at the cluster level following False Discovery Rate (FDR) correction.

For the HB specific analyses, percent signal change was extracted for left HB, right HB and the combined bilateral HB masks using Marsbar[33] and analysed with SPSS v15. Separate repeated-measures ANOVAS (rmANOVA) on the individual MID phases were run per ROI with Greenhouse-Geisser correction applied when applicable and corresponding significant post-hoc *t*-tests were Bonferroni-corrected (p_c).

2.6 Task-based functional connectivity of the habenula

To enable the assessment of functional connectivity between the left and right HB and the whole FOV at the group level, habenula seed ROIs, separate for left and right, were created by manually selecting each individual habenular voxel on the study-specific T1w template (in MNI space) using the software MRIcron[31]. Functional connectivity between the HB seed regions and the whole FOV was assessed using the generalized task-dependent psychophysiological interaction toolbox (gPPI, http://www.nitrc.org/projects/gppi), which calculates functional connectivity based on the deconvolved first eigenvariate of the seed time series. Functional connectivity was assessed while correcting for physiological variables and motion as described for the activation analyses and using the same statistical thresholds and one-sample *t*-tests on the contrasts between conditions.

3. Results

3.1 Behavioural

Behaviourally, participants responded correctly and on time to 47% of loss trials and 44% of neutral trials. Response times for correct trials were significantly faster for loss (M = 321.06, SD = 37.85 ms) than neutral (M = 361.11, SD = 40.55 ms) trials ($t(17) = 5.96$, $p < .001$). Concurrently, the performance-based response window duration was significantly shorter during loss (M = 360.97, SD = 46.08 ms) trials than neutral (M = 398.70, SD = 45.63 ms) trials ($t(17) = 5.40$, $p < .001$). For 17 participants, data were available to delineate the most common type of wrong responses, which were correct responses occurring outside the response window for loss (M = 91.99%, SD = 6.83) as well for neutral trials (M = 95.02%, SD = 6.16).

3.2 fMRI Whole FOV

Cluster level statistics and peak locations are provided in Table 1. Comparing the loss and neutral cues revealed significantly higher activation to loss cues in four clusters, two occipital, one in the left anterior insula, and one in the left caudate extending into the thalamus proper (see Figure 3a). No significant clusters were found for the reverse contrast.

Next, the activation differences between loss and neutral anticipation phases was assessed. While on the cluster level no significant differences remained following FDR correction, the right

inferior frontal gyrus was marginally more active during anticipation of loss than neutral trials (q_{FDR} = .056, z = 4.35, k_E = 761, MNI: 43, 40, 1). No significant clusters or peaks were detected for the reverse contrast.

To compare the neural activation patterns during the MID outcome phase, first, loss avoidance outcomes were compared to neutral outcomes. Whereas the reverse contrast did not reveal statistically significant clusters or peaks, enhanced activation for neutral over loss avoidance was found for two occipital clusters, one containing the bilateral lingual gyrus, the other the right middle and inferior occipital gyrus, shown in Figure 3b. Next, monetary loss outcomes were compared to neutral outcomes. Significantly higher activation to monetary loss than neutral outcomes was found for one cluster in the right middle temporal gyrus, see Figure 3c. In terms of increased neural activation to neutral than to monetary loss outcomes, three significant clusters were identified. We observed a decrease in activity to monetary loss in the left anterior insula, the right fusiform gyrus and in the right caudate. The caudate cluster originated in the right hemisphere, but contains bilateral caudate, bilateral putamen, bilateral NACC as well as subpeaks in the bilateral bed nucleus of the stria terminalis (BNST), see Figure 3d.

No significant differences emerged on peak or cluster level when comparing loss avoidance to monetary loss outcomes.

3.3 fMRI Habenula

The individual rmANOVAS on percent signal change during cue (right HB: $F(1,17) = .51$, p = .49, left HB: $F(1,17) = .01$, p = .94, bilateral HB: $F(1,17) = .19$, p = .67) and anticipatory phases (right HB: $F(1,17) = 2.80$, p = .11, left HB: $F(1,17) = .33$, p = .57, bilateral HB: $F(1,17) = 1.25$, p = .28) did not reveal differences between neutral and loss trials.

The rmANOVA on percent signal change during the outcome phase, divided into neutral, loss avoidance and monetary loss outcomes, revealed a main effect of outcome type (see Figure 4) for the right HB ($F(2,34) = 7.68$, $p_c < .01$), the left HB ($F(2,34) = 4.95$, $p_c < .05$) and bilateral HB ($F(2,34)$ =

7.17, $p_c < .01$) following Bonferroni-correction. Post-hoc comparisons indicated lower activity during loss avoidance outcomes (right HB: *M* = -.12, *SD* = .65; left HB: *M* = -.18, *SD* = .47, bilateral HB: *M* = -.15, *SD* = .50) as compared to neutral (right HB: *M* = .02, *SD* = .65; left HB: *M* = -.06, *SD* = .48, bilateral HB: *M* = -.02, *SD* = .52) outcomes for the right ($t(17) = 4.85$, $p_c < .001$), left ($t(17) = 3.48$, $p_c < .01$), and bilateral HB ($t(17) = 4.56$, $p_c < .001$).

Enhanced signal change was also observed during monetary loss (right HB: *M* = -.05, *SD* = .63; left HB: *M* = -.05, *SD* = .49, bilateral HB: *M* = -.05, *SD* = .51) as compared to loss avoidance outcomes for the left HB ($t(17) = 2.86$, $p_c < .05$), while not significant for the right HB ($t(17) = 1.53$, $p = .144$). The difference between loss avoidance and monetary loss in the bilateral HB did not remain significant following Bonferroni correction ($t(17) = 2.65$, $p_c = .05$). In other words, monetary loss outcomes were associated with significantly increased left habenula activity compared to avoiding loss or a potentially rewarding outcome.

The comparisons between neutral outcomes and monetary loss was not significant for right ($t(17) = 2.16$, $p_c = .14$), left ($t(17) = .11$, $p = .912$) or bilateral HB ($t(17) = .84$, $p = .41$).

3.4 Task-dependent functional connectivity with the Habenula

Comparing functional connectivity as indexed by gPPI during the cue phase across loss and neutral conditions did not reveal significant coupling differences between the left or right HB and other regions at $q_{FDR} < .05$.

When comparing loss anticipation to neutral trial anticipation, a significant positive slope represented the relationship between the right HB and the right hippocampus ($q_{FDR} < .05$, $z = 4.87$, $k_E = 510$, MNI: 31 -20 -12), see Figure 5a. Similarly, a significant positive slope was found for the relationship between the right HB and the subcallosal cingulate ($q_{FDR} < .05$, $z = 3.94$, $k_E = 514$, MNI: -2 13 -12), see Figure 5b.

When contrasting the functional connectivity across different outcome types, no significant differences in functional connectivity between the left and right HB seeds and other brain areas were found.

4. Discussion

In this first ultra-high field neuroimaging study on human HB function during loss avoidance, we show activity as expected of the HB given its upstream role in modulation of midbrain dopaminergic function[2]. We showed decreased HB activity, in both hemispheres, to loss avoidance outcomes (which are effectively acting as reward, relative to neutral and monetary loss). Behaviourally, response times were faster during loss avoidance than neutral trials indicating enhanced motivation and potentially salience of these trials. Greater left HB activity was also observed during monetary loss versus loss avoidance outcomes. During anticipation of loss relative to neutral outcomes, we found enhanced right HB functional connectivity with the subcallosal cingulate (SCA) and hippocampus.

In the whole FOV approach, even at ultra-high field (7T), HB-specific activation was not identifiable due to cluster size thresholding and multiple comparison corrections and larger smoothing kernels, further highlighting the need for an ROI approach to investigate HB functionality[8]. Single-cell LHB recordings have shown greater firing to the most negative outcome among alternatives and with inhibition to reward, especially at low predictability[1]. Similarly, human 3T fMRI studies showed enhanced HB activity to both aversive shock and when comparing loss relative to reward outcomes[13, 14]. Loss avoidance outcomes here behave similarly to reward outcomes: a deactivation of HB would be presumably associated with greater midbrain dopamine release. The monetary loss outcomes also show greater activity than loss avoidance outcomes which presumably would be associated with a cessation of midbrain dopamine activity[34].

Our primary whole-FOV findings were in the outcome phase, demonstrating that monetary loss relative to neutral outcomes was associated with deactivation of bilateral caudate, putamen,

BNST and NACC, and the left anterior insula. The NACC, caudate and putamen have previously been shown to be differentially responsive to the anticipation of neutral and monetary loss trials with putaminal activity further reported in the MID outcome phase[24]. The NACC and anterior insular activity have previously been associated with tracking negative prediction error[35, 36].

Loss cues in the whole-FOV analysis were associated with activity in regions implicated in loss and value representation with greater predominantly left-sided activity in the anterior insula and caudate. The loss cue, although associated with the opportunity to avoid losing, predicted an increased chance on losing than the neutral cue.

In contrast to previous meta-analyses which have shown similar activations during the anticipatory phase of reward and loss trials in the MID task[37-39], we did not observe any differential activity in the anticipation phase in the whole-FOV analyses. However, we found differences in functional connectivity with greater connectivity between the right HB and hippocampal and subcallosal cingulate. The hippocampal involvement likely reflects underlying memory-related processing[40] and has been shown in the anticipatory phase in the MID task[41]. Hippocampal activity has been previously shown to linearly scale with loss magnitudes[42]. Rodent lesion studies suggest that the hippocampal complex is especially relevant at learning the initial matching between cues and outcomes as well as during memory processing relating to non-specific reward expectancy[40]. The role of the hippocampus may relate to information transfer between the anticipatory cue and associated potential outcomes.

The SCA is a projection target for midbrain dopaminergic neurons modulated by HB activity[2, 43]. Resting-state 3T functional connectivity of the human HB has previously identified enhanced functional coupling between the HB and the SCA[44]. Similarly, a probabilistic Pavlovian learning paradigm with monetary rewards, losses and electric shocks as punishment showed a non-significant increase in functional coupling between the right HB and Brodmann Area 25, with increasing motivational value of the punishment-related conditioned stimulus[14].

Our findings might be particularly relevant in the context of major depression. Depression is associated with abnormal subcallosal cingulate activity and connectivity patterns[45-48] and positive effects on depressive symptoms were reported with deep brain stimulation (DBS) targeting the SCA[49]. A similar remission of depressive symptoms has been seen in a case study following DBS to the lateral HB in an MDD patient[3] and in a patient with bipolar disorder with refractory depression[4].

The current study is not without limitations. In light of the intended functional connectivity and FOV brain analyses in addition to the HB ROI analyses, we attempted to include core MID task-related structures such as the hippocampus and subcallosal cingulate in the assessed FOV. This, meant that we could not centre the FOV over the HB, which enabled one participant to move their HB out of the FOV during the scan and corresponding data was subsequently discarded from all analyses. We were also unable to distinguish between lateral and medial HB. Further, HB ROIs were not independently confirmed by other raters. We address this issue by presenting results of ROIs created two ways: based on 7T image contrast and using the geometric method[8]. Finally, while we controlled for known confounds, such as cardiac rhythm and respiratory rate[50], recent evidence also hints towards a possible effect of circadian rhythm[51]. Given the inconsistencies regarding laterality findings of HB function across previous research, we chose to analyse left, right and bilateral HB separately without directly assessing lateralization. Of note, the majority of trials leading to negative feedback in the loss condition were correct responses occurring outside the response window. While staircase procedures, hence adjusting the allowed response window, are commonly utilized in MID tasks (e.g., [25-28]), incorrect button presses occurred at a low rate. As such we suggest that investigations aiming to specifically delineate the HB response to behavioural errors utilize task designs evoking higher proportions of incorrect button presses.

In summary, we demonstrate that HB activity differentiates between monetary loss avoidance, monetary loss and neutral outcomes for the first time in an ultra-high field (7T) subcortical task-based fMRI study. Our findings thus converge with proposed HB function in rodent studies and extend previous observations in human imaging studies at 3T. The HB appears to be a

critical structure particularly in depressive disorders and has been implicated as a potential key node in the anti-depressive mechanism of action of ketamine[7]. Further studies using task-based fMRI at 7T to investigate the role of the HB in depression and the effects of ketamine are warranted.

Funding and Disclosure

The authors declare no conflict of interest. This research was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). VV is funded by a Medical Research Council Senior Fellowship (MR/P008747/1). CLN is funded by the Medical Research Council (Grant Reference: SUAG/043 G101400). CTR is funded by a Sir Henry Dale Fellowship from the Welcome Trust and the Royal Society [098436/Z/12/B]. The 7T MRI scanner was funded by the University of Cambridge, the Medical Research Council (MR/M008983/1), and the NIHR Cambridge Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

No part of the study procedures and analyses was pre-registered prior to the research being conducted. We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analyses, all manipulations, and all measures in the study. Study data, digital study materials, and analysis code are available via https://doi.org/10.17863/CAM.66358.

Acknowledgments

We would like to thank the Wolfson Brain Imaging Centre and their staff for their technical expertise and help in piloting and setting up sequences and scanning participants.

Author Contributions

KW supported all aspects of the project. SGK and CLN supported the data analyses. CR and CTR supported the data acquisition. VV acquired the funding and supported study design and analyses. All authors supported the drafting of the manuscript and agree to this publication.

References

1. Matsumoto M, Hikosaka O. Representation of negative motivational value in the primate lateral habenula. *Nature neuroscience* 2009; **12**(1)**:** 77-84.

2. Matsumoto M, Hikosaka O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 2007; **447**(7148)**:** 1111-1115.

3. Sartorius A, Kiening KL, Kirsch P, von Gall CC, Haberkorn U, Unterberg AW *et al.* Remission of major depression under deep brain stimulation of the lateral habenula in a therapy-refractory patient. *Biological psychiatry* 2010; **67**(2)**:** e9-e11.

4. Zhang C, Kim SG, Li D, Zhang Y, Li Y, Husch A *et al.* Habenula deep brain stimulation for refractory bipolar disorder. *Brain stimulation* 2019; **12**(5)**:** 1298-1300.

5. Yang Y, Wang H, Hu J, Hu H. Lateral habenula in the pathophysiology of depression. *Current opinion in neurobiology* 2018; **48:** 90-96.

6. Lawson R, Nord C, Seymour B, Thomas D, Dayan P, Pilling S *et al.* Disrupted habenula function in major depression. *Molecular psychiatry* 2017; **22**(2)**:** 202-208.

7. Yang Y, Cui Y, Sang K, Dong Y, Ni Z, Ma S *et al.* Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. *Nature* 2018; **554**(7692)**:** 317-322.

8. Lawson RP, Drevets WC, Roiser JP. Defining the habenula in human neuroimaging studies. *NeuroImage* 2013; **64:** 722-727.

9. Akagi K, Powell EW. Differential projections of habenular nuclei. *Journal of Comparative Neurology* 1968; **132**(2)**:** 263-273.

10. Namboodiri VM, Rodriguez-Romaguera J, Stuber GD. The habenula. *Current biology : CB* 2016; **26**(19)**:** R873-R877.

11. Epstein EL, Hurley RA, Taber KH. The Habenula's Role in Adaptive Behaviors: Contributions From Neuroimaging. *The Journal of neuropsychiatry and clinical neurosciences* 2018; **30**(1)**:** A4-4.

12. Hennigan K, D'Ardenne K, McClure SM. Distinct midbrain and habenula pathways are involved in processing aversive events in humans. *The Journal of Neuroscience* 2015; **35**(1)**:** 198-208.

13. Furman DJ, Gotlib IH. Habenula responses to potential and actual loss in major depression: preliminary evidence for lateralized dysfunction. *Social cognitive and affective neuroscience* 2016; **11**(5)**:** 843-851.

14. Lawson RP, Seymour B, Loh E, Lutti A, Dolan RJ, Dayan P *et al.* The habenula encodes negative motivational value associated with primary punishment in humans. *Proceedings of the National Academy of Sciences* 2014; **111**(32)**:** 11858-11863.

15. Liu WH, Valton V, Wang LZ, Zhu YH, Roiser JP. Association between habenula dysfunction and motivational symptoms in unmedicated major depressive disorder. *Social cognitive and affective neuroscience* 2017; **12**(9)**:** 1520-1533.

16. Garrison J, Erdeniz B, Done J. Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews* 2013; **37**(7)**:** 1297-1310.

17. Flannery JS, Riedel MC, Poudel R, Laird AR, Ross TJ, Salmeron BJ *et al.* Habenular and striatal activity during performance feedback are differentially linked with state-like and trait-like aspects of tobacco use disorder. *Science Advances* 2019; **5**(10)**:** eaax2084.

18. Bjork JM, Knutson B, Hommer DW. Incentive-elicited striatal activation in adolescent children of alcoholics. *Addiction* 2008; **103**(8)**:** 1308-1319.

19. Scheres A, Milham MP, Knutson B, Castellanos FX. Ventral striatal hyporesponsiveness during reward anticipation in attention-deficit/hyperactivity disorder. *Biological psychiatry* 2007; **61**(5)**:** 720-724.

20. Filbey FM, Dunlop J, Myers US. Neural effects of positive and negative incentives during marijuana withdrawal. *PloS one* 2013; **8**(5)**:** e61470.

21. Stoy M, Schlagenhauf F, Schlochtermeier L, Wrase J, Knutson B, Lehmkuhl U *et al.* Reward processing in male adults with childhood ADHD—a comparison between drug-naive and methylphenidate-treated subjects. *Psychopharmacology* 2011; **215**(3)**:** 467-481.

22. Beck A, Schlagenhauf F, Wüstenberg T, Hein J, Kienast T, Kahnt T *et al.* Ventral striatal activation during reward anticipation correlates with impulsivity in alcoholics. *Biological psychiatry* 2009; **66**(8)**:** 734-742.

23. Romanczuk-Seiferth N, Koehler S, Dreesen C, Wustenberg T, Heinz A. Pathological gambling and alcohol dependence: neural disturbances in reward and loss avoidance processing. *Addiction biology* 2015; **20**(3)**:** 557-569.

24. Knutson B, Fong GW, Adams CM, Varner JL, Hommer D. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* 2001; **12**(17)**:** 3683-3687.

25. Nestor LJ, Murphy A, McGonigle J, Orban C, Reed L, Taylor E *et al.* Acute naltrexone does not remediate fronto-striatal disturbances in alcoholic and alcoholic polysubstance-dependent populations during a monetary incentive delay task. *Addiction biology* 2017; **22**(6)**:** 1576-1589.

26. Mei S, Li Q, Liu X, Zheng Y. Monetary incentives modulate feedback-related brain activity. *Scientific reports* 2018; **8**(1)**:** 1-13.

27. Miller EM, Shankar MU, Knutson B, McClure SM. Dissociating motivation from reward in human striatal activity. *Journal of cognitive neuroscience* 2014; **26**(5)**:** 1075-1084.

28. Zhang Y, Li Q, Wang Z, Liu X, Zheng Y. Temporal dynamics of reward anticipation in the human brain. *Biological psychology* 2017; **128:** 89-97.

29. Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele P-F, Gruetter R. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage* 2010; **49**(2)**:** 1271-1281.

30. Ashburner J, Friston KJ. Unified segmentation. *NeuroImage* 2005; **26**(3)**:** 839-851.

31. Rorden C, Brett M. Stereotaxic display of brain lesions. *Behavioural neurology* 2000; **12**(4)**:** 191-200.

32. Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinzle J, Iglesias S *et al.* The PhysIO toolbox for modeling physiological noise in fMRI data. *Journal of neuroscience methods* 2017; **276:** 56-72.

33. Brett M, Anton J-L, Valabregue R, Poline J-B. Region of interest analysis using the MarsBar toolbox for SPM 99. *NeuroImage* 2002; **16**(2)**:** S497.

34. Fiorillo CD, Tobler PN, Schultz W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 2003; **299**(5614)**:** 1898-1902.

35. Harrison NA, Voon V, Cercignani M, Cooper EA, Pessiglione M, Critchley HD. A Neurocomputational Account of How Inflammation Enhances Sensitivity to Punishments Versus Rewards. *Biological psychiatry* 2016; **80**(1)**:** 73-81.

36. Voon V, Pessiglione M, Brezing C, Gallea C, Fernandez HH, Dolan RJ *et al.* Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron* 2010; **65**(1)**:** 135-142.

37. Dugre JR, Dumais A, Bitar N, Potvin S. Loss anticipation and outcome during the Monetary Incentive Delay Task: a neuroimaging systematic review and meta-analysis. *PeerJ* 2018; **6:** e4749.

38. Oldham S, Murawski C, Fornito A, Youssef G, Yucel M, Lorenzetti V. The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human brain mapping* 2018; **39**(8)**:** 3398-3418.

39. Wilson RP, Colizzi M, Bossong MG, Allen P, Kempton M, Bhattacharyya S. The Neural Substrate of Reward Anticipation in Health: A Meta-Analysis of fMRI Findings in the Monetary Incentive Delay Task. *Neuropsychology review* 2018; **28**(4)**:** 496-506.

40. Savage LM, Buzzetti RA, Ramirez DR. The effects of hippocampal lesions on learning, memory, and reward expectancies. *Neurobiology of learning and memory* 2004; **82**(2)**:** 109-119.

41. Patel KT, Stevens MC, Meda SA, Muska C, Thomas AD, Potenza MN *et al.* Robust changes in reward circuitry during reward loss in current and former cocaine users during performance of a monetary incentive delay task. *Biological psychiatry* 2013; **74**(7)**:** 529-537.

42. Hahn T, Dresler T, Plichta MM, Ehlis AC, Ernst LH, Markulin F *et al.* Functional amygdala-hippocampus connectivity during anticipation of aversive events is associated with Gray's trait "sensitivity to punishment". *Biological psychiatry* 2010; **68**(5)**:** 459-464.

43. McInerney SJ, McNeely HE, Geraci J, Giacobbe P, Rizvi SJ, Ceniti AK *et al.* Neurocognitive predictors of response in treatment resistant depression to subcallosal cingulate gyrus deep brain stimulation. *Frontiers in human neuroscience* 2017; **11:** 74.

44. Erpelding N, Sava S, Simons LE, Lebel A, Serrano P, Becerra L *et al.* Habenula functional resting-state connectivity in pediatric CRPS. *Journal of neurophysiology* 2014; **111**(2)**:** 239-247.

45. Roiser JP, Levy J, Fromm SJ, Nugent AC, Talagala SL, Hasler G *et al.* The effects of tryptophan depletion on neural responses to emotional words in remitted depression. *Biological psychiatry* 2009; **66**(5)**:** 441-450.

46. Tozzi L, Doolin K, Farrel C, Joseph S, O'Keane V, Frodl T. Functional magnetic resonance imaging correlates of emotion recognition and voluntary attentional regulation in depression: A generalized psycho-physiological interaction study. *Journal of affective disorders* 2017; **208:** 535-544.

47. Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H *et al.* Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological psychiatry* 2007; **62**(5)**:** 429-437.

48. Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C *et al.* Deep brain stimulation for treatment-resistant depression. *Neuron* 2005; **45**(5)**:** 651-660.

49. Holtzheimer PE, Kelley ME, Gross RE, Filkowski MM, Garlow SJ, Barrocas A *et al.* Subcallosal cingulate deep brain stimulation for treatment-resistant unipolar and bipolar depression. *Archives of general psychiatry* 2012; **69**(2)**:** 150-158.

50. Hutton C, Josephs O, Stadler J, Featherstone E, Reid A, Speck O *et al.* The impact of physiological noise correction on fMRI at 7 T. *NeuroImage* 2011; **57**(1)**:** 101-112.

51. Kaiser C, Kaufmann C, Leutritz T, Arnold YL, Speck O, Ullsperger M. The human habenula is responsive to changes in luminance and circadian rhythm. *NeuroImage* 2019; **189:** 581-588.