Non-invasive Ultrasound Deep Neuromodulation of the Human Nucleus Accumbens Increases Win-Stay Behaviour

Siti N. Yaakub[1,2], Nadège Bault[1,2], Mathilde Lojkiewiez[1,2], Elouan Bellec[1,2], Jamie Roberts[5], Noah S. Philip[3,4], Matthew F.S. Rushworth[6,7*], Elsa F. Fouragnan[1,2*]

[1]School of Psychology, Faculty of Health, University of Plymouth, Plymouth, United Kingdom

[2]Brain Research and Imaging Centre, Faculty of Health, University of Plymouth, Plymouth, United Kingdom

[3]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA

[4]Center for Neurorestoration and Neurotechnology, VA Providence Healthcare System, Providence, Rhode Island, USA

[5]Department of Clinical Measurement and Innovation, University Hospitals Plymouth NHS Trust, Plymouth, UK

[6]Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK

[7]Department of Experimental Psychology, University of Oxford, Oxford, UK

* People contributed equally to this work

Correspondence and material requests should be addressed to Elsa F. Fouragnan (elsa.fouragnan@plymouth.ac.uk).

Summary

Precisely neuromodulating deep brain regions in humans could bring transformative advancements in both cognitive neuroscience and brain disorder treatment. In a within subject experiment, twenty-six healthy adults underwent a series of transcranial ultrasound stimulation procedures, including stimulation of the nucleus accumbens, the dorsal anterior cingulate cortex, or no sonication. Results revealed that ultrasound stimulation of the nucleus accumbens (NAcc) induced changes in reward-related behaviours, including in the tendency to stick with winning choices, the rate of learning specifically from positive feedback, and the rate of repeating rewarded choices. Functional brain scans showed corresponding neural changes in response to reward expectations in targeted and interconnected brain areas. These findings demonstrate the causal role of the human NAcc in learning from positive feedback, as well as the feasibility of using non-invasive neuromodulation deep in the human brain to modulate learning and decision making both as a research tool and as a potential component of future treatments for disorders involving reward sensitivity.

Abstract

Current methods for directly modulating activity in deep human brain structures are limited and invasive. Here, we demonstrate modulation of activity in human deep brain structures using non-invasive transcranial ultrasound stimulation (TUS) and show it exerts selective effects on learning and decision making. 26 healthy adults participated in a within-subject repeated TUS–functional magnetic resonance imaging (fMRI) experiment with three sonication conditions. Participants performed a probabilistic learning task while undergoing fMRI scanning after counterbalanced 80 s 5 Hz repetitive TUS of the nucleus accumbens (NAcc), dorsal anterior cingulate cortex (dACC), or Sham. TUS-NAcc altered parametric BOLD response to reward expectation in NAcc and adjacent areas. Moreover, it induced changes specific to reward-related behaviours including in the use of win–stay strategy, learning rate following rewards, learning curve for choices of the rewarding option, and rate of repetition for rewarded choices. The results demonstrate TUS's utility for human non-invasive deep-brain neuromodulation and the causal role of the human NAcc in learning from positive outcomes.

Introduction

Brain disorders—including neurological, psychiatric, and developmental—exact a heavy toll not only on individuals, profoundly affecting cognitive abilities and emotional well-being, but also on entire communities, hindering societal dynamics, workforce productivity, and straining resources [1]. Neuromodulation holds the potential to complement or even surpass current treatments for brain disorders by offering a targeted, nonpharmacological approach that directly modulates neural activity, providing a more personalized and effective therapeutic strategy [2,3]. However, existing neuromodulation methods have limitations that restrict their effectiveness for treating brain disorders when potential promising targets often lie in circumscribed regions deep in the brain [4] . They are either invasive and require surgery (deep brain stimulation), have modest spatial resolution but are limited to the dorsolateral surface (transcranial magnetic stimulation), or non-specifically modulate a broad area of tissue (transcranial direct current stimulation).

An alternative approach involves harnessing ultrasound waves, typically employed in diagnostics and therapeutic interventions, to precisely target specific brain regions that are responsive to acoustic energy [5]. The method, called transcranial ultrasound stimulation (TUS), focuses ultrasound beams through the skull onto the brain, altering neural activity noninvasively [6]. TUS can safely induce neural activity changes in deep brain regions with millimetric resolution and modulate discrete cell types within specific brain regions, and it can do so without affecting overlying brain areas [7].

Repetitive TUS, lasting tens of seconds to minutes, produces neural changes that outlast the stimulation period itself, with some effects lasting hours and resembling early phase neuroplasticity [8,9]. Recent proof-of-concept studies using repetitive TUS in non-human primates show changes in neural activity when TUS is targeted at deep cortical and subcortical regions [10–12]. These studies have been supplemented by demonstrating that TUS also induces changes in cognition and behaviour [11,13–17].

In parallel, when careful precautions are taken to limit the transmission loss caused by the skull [18,19], and with experimental designs that prevent auditory confounds (e.g. “offline” TUS protocols), some emerging studies in humans have demonstrated that TUS can effectively and precisely modulate activity and neurochemistry in deep parts of the cortex, such as the posterior cingulate cortex (PCC), while participants are at rest [20]. However, it remains unclear whether TUS can also be used to produce targeted changes in behaviour and neural activity by causally manipulating deep subcortical structures while participants are engaged in higher cognitive processes such as decision making and learning which is our focus in the current

study. If this is possible, then not only does it open new avenues for testing causal hypotheses regarding human brain function, but it also raises the possibility of refined clinical intervention strategies and precision therapeutic applications.

Here we focus on the nucleus accumbens (NAcc), a deep subcortical region implicated in reward-guided learning and memory in humans and other animals [21–24]. It is a key target of amygdala and mesolimbic dopaminergic projections [25,26]. Such inputs allow the NAcc to guide behaviour based on the reward outcomes expected after choices are made, and to facilitate learning and adaptation by signalling disparities between predicted rewards and actual rewards. Midbrain dopamine neurons projecting onto the NAcc carry both anticipated reward and prediction error signals which can be described with classical reinforcement learning models [27,28] [27,29].

In humans, functional magnetic resonance imaging (fMRI) studies investigating the blood oxygenation-level dependant (BOLD) response to expected reward and prediction error align with this hypothesis, with evidence of model-based prediction error activity in midbrain regions [30] and the NAcc [22,24,31–36]. Related signals are also found within a larger distributed reward network encompassing regions such as the ventromedial prefrontal cortex (vmPFC), the lateral orbitofrontal cortex (lOFC), the dorsal anterior cingulate cortex (dACC), and the PCC.

Building on our own and others' previous work in humans and macaques, we sought to test whether TUS could be used to manipulate activity in the human NAcc. While we use techniques that we have previously used to measure changes in neural activity, an important aspect of our present investigation was to ascertain whether neural changes were associated with change at a behavioural level. We were, therefore, also guided by studies that have carried out causal manipulations of the NAcc in another primate species, macaques [37,38,38–42]. In aggregate, these studies emphasize the importance of stochastic rather than deterministic reward schedules for investigating the NAcc and that NAcc interventions in primates specifically affect reward-guided, as opposed to loss-guided, aspects of behaviour. We therefore used a related task in the current investigation and took special care to assess the degree to which behavioural changes were specific to reward-guided aspects of learning and decision making. NAcc lesions in macaques disrupt reward-guided learning and decision making and TUS in previous macaque studies have disrupted the behaviours that have been investigated. In the current study, however, we took care to employ an especially cautious set of TUS parameters when testing human participants. We were therefore open to the possibility that changes in one of these parameters might lead to a facilitative or a disruptive effect even

while retaining the expectation that any effects would be specific to reward-guided learning and decision making.

A total of 26 healthy participants took part in a within-subject repeated TUS–fMRI experiment with three sonication conditions. Following 5 Hz repetitive TUS applied in a counterbalanced manner to the NAcc (TUS-NAcc), dACC (TUS-dACC), or not at all (Sham), participants engaged in a probabilistic reversal learning task inside an MRI scanner. TUS-NAcc resulted in specific alterations in reward-related behaviours, such as changes in the win–stay strategy, learning rates following rewards, the learning curve for the rewarding option, and choice repetition rates after rewards. TUS elicited neural increases in the parametric BOLD response related to reward expectation and delivery in the sonicated and adjacent areas. BOLD changes linked to reward signals in interconnected areas, even when distant, were, also observed.

This study demonstrates that TUS applied to the human NAcc induces specific alterations in reward-related behaviours and corresponding neural changes, showcasing its capacity to causally manipulate deep subcortical structures and influence decision-making processes, thereby highlighting its potential as a promising avenue for therapeutic interventions targeting brain disorders, with the possibility of offering a highly precise and potentially more effective treatment strategy.

Results

TUS approach and acoustic simulations

Figure 1 summarises the TUS approach, personalised planning, and results of acoustic simulation of TUS application. Targeting of the left NAcc and dACC was determined based on Montreal Neurological Institute (MNI) coordinates, individually adjusted using T1-weighted MRI scans during target and transducer placement planning and acoustic simulation (Fig. 1a-c). Repetitive TUS comprised a 5 Hz-patterned protocol with a 10% duty cycle applied for 80 seconds. The spatial peak pulse average intensity (I_{SPPA}) in water was maintained at ~30W/cm^2 across participants, previously validated in a hydrophone tank [20,43]. Sham TUS was administered similarly to active TUS, except that no stimulation was delivered. Instead, participants heard a sound mimicking the TUS protocol sound, played via bone conduction headphones (Fig 1b). Post-study verbal prompts revealed no discernible difference in perception between the three sessions among participants. Acoustic simulation for each participant was carefully planned to ensure on-target pressure and safety (an example participant is illustrated in Fig. 1c). Figure 1d displays the averaged pressure maps across

individuals for each target region. The acoustic simulation parameters and output for all study participants can be found in Supplementary Tables 1 and 2. Both the NAcc and dACC were associated with similar mechanical indices, I_{SPPA} at focus (derated) and occurrence of side effects (Fig. 1e; see Methods section under TUS protocol and procedure for more details on safety).

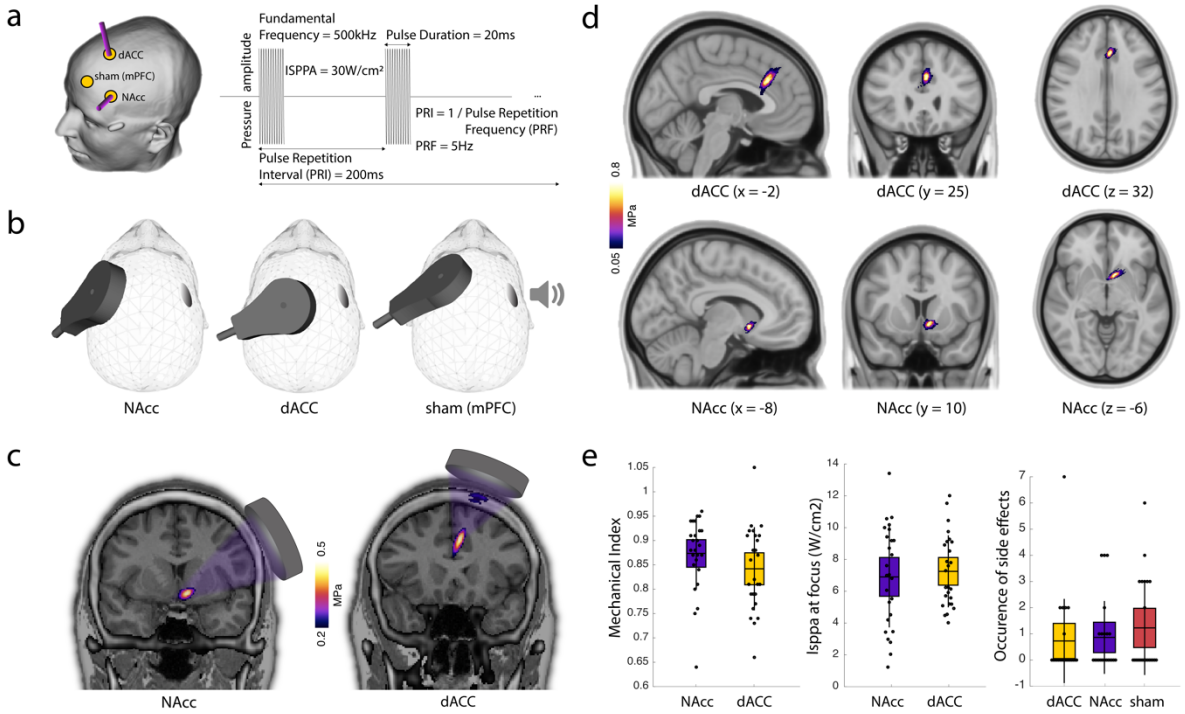

Fig. 1. TUS intervention. a Target trajectories for neuronavigation and TUS protocol. **b** Visualisation of the positioning of the transducer on the head and of the bone conductive headphones. **c** Example of an individual subject's personalised TUS planning with pressure field displayed for the upper half of the full width half maximum (-6dB). The pseudo-CT (thresholded at 300 HU) is also displayed in grey, from which Hounsfield units are mapped onto acoustic properties of the skull. **d** Averaged pressure maps across all participants restricted to each focal volume (as defined by -6dB). The top row represents the average pressure map (in NIH fire colour from FSLeyes) across all individuals for the dACC (n=26) while the bottom row represents the average pressure map for the NAcc (n=26). **e** Average and standard error of (left panel) mechanical index, (middle panel) I_{SPPA} and (right panel) occurrence of side effects after TUS-fMRI in the two TUS targets. Data from each individual participant are presented as circles.

As TUS in humans is still relatively new compared to TUS in non-human primates, it is worth highlighting some differences between the TUS protocols employed in the two species, namely (1) the ultrasound intensity employed here is considerably lower than the level employed in

previous macaque studies where typical intensities in the brain, at focus, has ranged between 24W/cm^2–64W/cm^2, (2) the pulse repetition frequency and duty cycle are different, and (3) there might be a loss in coupling at the interface between the transducer and the skin that is unaccounted for in humans; by contrast, animals are usually shaved [11–14]. While the intensities used in macaque studies have been shown to lead to no tissue damage, we took care to remain within the guidelines for human ultrasound exposure as defined by ITRUSST [44].

Study design

A total of 26 healthy individuals participated in a within-subject repeated measures design study with three visits after an initial screening and MRI acquisition used for planning TUS (session 1). Each of the subsequent three visits included TUS targeted at the NAcc (TUS-NAcc), dACC (TUS-dACC), or a no sonication (Sham) condition. Visits were spaced at least one week apart and administered in a counterbalanced order across participants to control for any potential order effects (sessions 2–4; Fig. 2a,c). During each of the three visits, participants first received TUS (or Sham) and were then placed in an MRI scanner and engaged in a probabilistic reversal learning task. Recordings of neural activity only begun approximately ~10 min after the end of TUS application when any potential auditory or somatosensory effects of stimulation were dissipated. We therefore refer to this stimulation protocol as an “offline” protocol.

Four blocks of the task were presented on average ~15, 28, 35, and 48 min after TUS (Fig. 2a,b). In each trial of the task, participants were presented with a pair of abstract symbols drawn from a larger pool of three symbols and, through feedback, learned to choose the symbol with the highest probability of reward [14,31,32]. We employed the same task in previous work in humans [32,45] and macaques [13,14]. This has a few advantages. Because the best option is not presented on every trial, participants may have to make different choices on different trials based on which options are available; the task cannot be performed by simply repeating choices. Moreover, this manipulation enables us to achieve a more even distribution of positive and negative outcomes. Essentially, it forced participants to choose between the two least rewarding symbols in every third trial, even after they have learnt the cue–reward contingencies. After 25 trials, the high reward probability was assigned to a different symbol, prompting participants to adapt to the new learning phase (i.e., a ‘reversal’ in reward contingencies occurred). Because the choice–reward associations were stochastic as opposed to deterministic, the task is similar to one previously found to be sensitive to NAcc intervention in macaques [42].

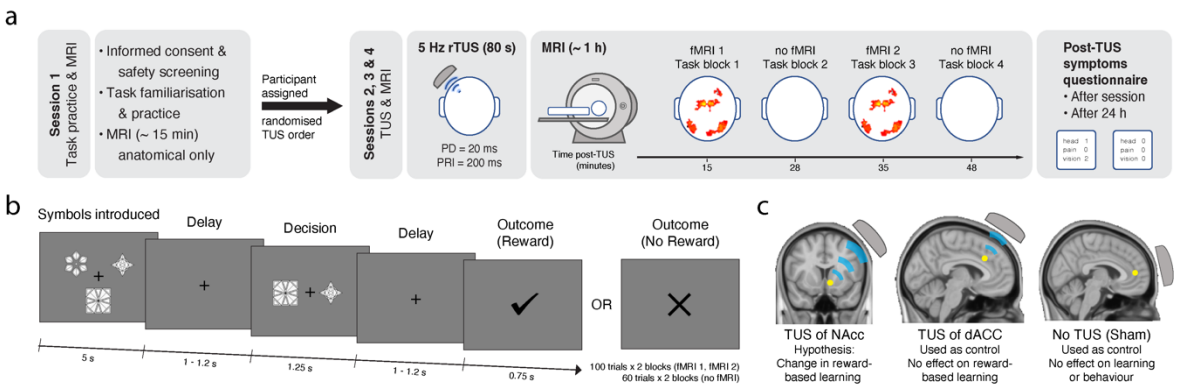

Fig. 2. Study design. a A total of 26 healthy participants attended four study sessions each. During the first session, participants practiced a short version of the reversal learning task and underwent a series of anatomical MRI scans. Participants were then assigned a randomised TUS condition order for the subsequent three TUS and MRI sessions (sessions 2, 3, and 4). During these sessions, participants first underwent 80 seconds of 5 Hz repetitive TUS (pulse duration (PD) = 20 ms and pulse repetition interval (PRI) = 200 ms) targeted at the region determined by the TUS condition order. This was immediately followed by approximately 1 hour of MRI scans in the sequence shown. At the end of each session, participants filled in a questionnaire asking about any possible symptoms related to TUS. The questionnaire was repeated at least 24 hours after each session. **b** Schematic representation of the probabilistic reversal learning task. At the start of each task block, participants were shown the three abstract symbols that would appear in that block. On each trial of the task, participants would see two of the three symbols and were required to select, by pressing one of two buttons, the symbol that was most likely to lead to a reward. After a short delay, they were presented with either a tick, indicating a reward, or a cross, indicating no reward. Participants were presented with 320 trials in total divided into four task blocks in the scanner: two blocks of 100 trials each (during the fMRI scans) and two blocks of 60 trials each (during which further MR scans were acquired but not used in this study). **c** Illustration of the three TUS conditions showing the location of each target (yellow circle) overlaid on the MNI template brain. TUS was either applied to the NAcc or dACC. During the Sham TUS condition, the transducer was positioned over medial frontal cortex, but no TUS was delivered.

Behavioural findings

During the task, participants learnt to choose the symbol with the highest probability of reward and avoided the other two symbols. Overall, participants' responses were probabilistic and aligned with the principles of a reinforcement learning mechanism. Model comparison revealed that a two learning rates model (2LRs)—one for learning from rewarded outcomes and one for learning from non-rewarded outcomes—performed better than a simple Rescorla–Wagner (RW) model at fitting behaviours (BIC_{RW} = 385.78, BIC_{2LRs} = 293.48). We did not observe any statistically significant difference between TUS conditions in terms of overall performance (analysis of variance, ANOVA: $F_{2,75}$ = 0.06, p = 0.93) or reaction times (ANOVA: $F_{2,75}$ = 0.38, p = 0.68) in the task. This was also true when the high probability symbol was on the screen

(ANOVA: performance: $F_{2,75}$ = 0.20, p = 0.81; RT: $F_{2,75}$ =0.21, p = 0.8; see all ANOVA results in Supplementary Table 3).

One of the main goals of this study was to determine whether TUS of the NAcc can induce specific changes in reward-related behaviours and whether this can be observed in the hour following TUS. For this, we focused on win–stay behaviours and ran a regression analysis across blocks, participants, and conditions (see Methods, Analysis A). The regression weights were then subjected to an ANOVA to check whether at least one condition was different to the others and post-hoc paired two-sided t-tests were performed, which are all fully reported in Supplementary Table 4. Our main analysis revealed a main effect of condition (ANOVA: $F_{2,75}$ = 3.3, p = 0.042). Post-hoc t-tests revealed a stronger relationship between reward and subsequent win–stay behaviour after TUS-NAcc compared to Sham (t_{25} = -2.75, p = 0.011 Bonferroni corrected, Cohen D = 0.53, see full results in Supplementary Table 4) and compared to TUS-dACC (t_{25} = -3.52, p = 0.001 Bonferroni corrected, Cohen D = 0.69). There was no significant difference between TUS-dACC and Sham (t_{25} = -0.57, p = 0.573, Cohen D = 0.11). This is presented in Figure 3a.

To test the temporal dynamics of TUS effects post-sonication, we then analysed whether the difference between TUS-NAcc and Sham was significantly different to 0 for each of the four blocks. These blocks happened at approximately 15-, 28-, 35-, and 48-minutes post TUS (averaged across participants and conditions). This analysis revealed that the effects of TUS-NAcc compared to Sham was statistically significant at 28- and 35-minutes post TUS (Fig. 3b left panel), but not at 15- and 48-minutes post TUS (see post-hoc tests in Supplementary Table 4). There was no statistically significant difference between dACC and Sham at any of the time points (Fig. 3b right panel; see post-hoc tests in Supplementary Table 4).

Having found an increase in win–stay behaviours indicating a potential difference in learning strategies, we then looked at learning curves across trials during a reversal period in the two blocks that were significant. We computed the rate of choice of the high probability option and found that participants were more likely to select the high probability option at the end of a reversal period after TUS-NAcc, compared to TUS-dACC and Sham (mixed-effect model: t_{5924} = -2.74, p = 0.006; Fig. 3c; Supplementary Table 5). This is particularly true for the first reversal of the block (t_{1478} = -3.97, p = 0.0007; Fig. 3d; Supplementary Table 5).

To look at the influence of reward history over a longer period than is captured by the one-trial back-based win-stay heuristic, we fitted the behavioural data to RL models and then repeated all previous analyses with our best fitting reinforcement learning model estimates. We tested the impact of prediction error (a weighted average of reward history) on subsequent win–stay

choices (Fig. 3e) and how the expected value of the high probability option compares across conditions. This model-based series of analyses suggested similar conclusions. Namely, we found a stronger relationship between PE and subsequent win–stay choices after TUS of the NAcc compared to Sham and dACC (ANOVA: $F_{2,75} = 3.33$, $p = 0.041$; see full ANOVA and post hoc t-tests in Supplementary Table 6). This relationship also peaked around 28–35 min after TUS (Supplementary Table 6). Finally, the expected value associated with the high probability option was significantly higher after TUS-NAcc compared to the other conditions (Supplementary Table 7), and this was also true for the first reversal of the block (Fig. 3f–h).

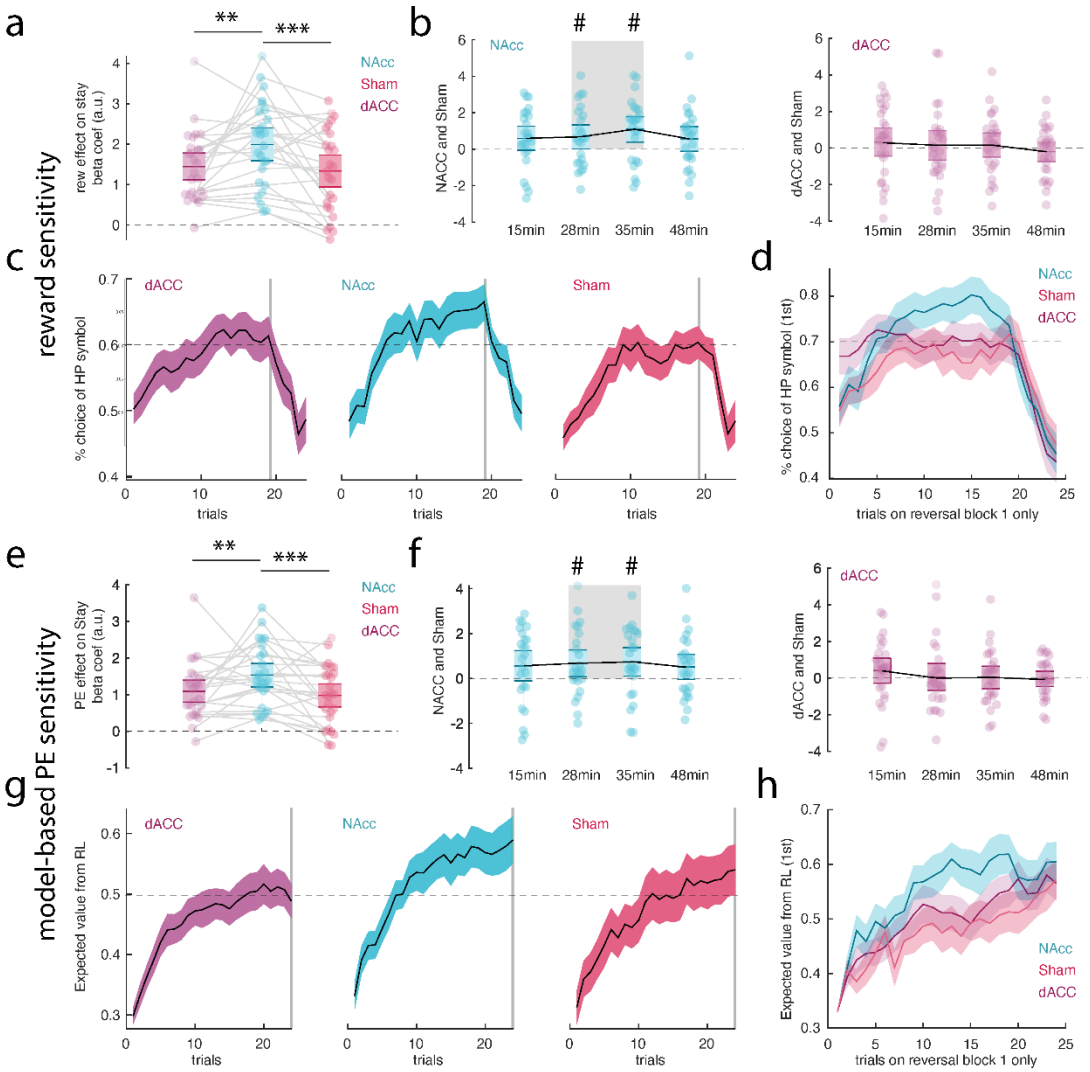

Fig. 3. Behavioural and model results. a Win–stay analyses revealed an increase in the relationship between reward and subsequent win–stay behaviours after NAcc TUS. **b** Temporally, the NAcc TUS-induced reward-related changes were prominent in the middle of the approximately hour-long post-TUS period. **c** Learning curves associated with the high probability option. **d** Same as **c** but showing only the first reversal block. **e** Same as **a** but

with PE instead of reward. **f** Temporal effects of PE-related TUS changes. **g** Increase in the expected value of the high probability option. **h** Same as **g** but for block 1.

Having established that TUS-NAcc impacted learning curves associated with the high probability option, we tested whether overall, participants exhibited higher learning rates after TUS-NAcc estimated from the RL model. This is indeed what we found. The learning rates associated with positive feedback were higher after TUS-NAcc compared to dACC and Sham (ANOVA: $F_{2,75} = 5.68$, $p = 0.005$; post hoc t-tests: TUS-dACC vs. TUS-NAcc: $t_{25} = -2.46$, $p = 0.021$; Cohen D = 0.48, TUS-NAcc vs. Sham: $t_{25} = -3$, $p = 0.006$; Cohen D = 0.59; TUS-dACC vs. Sham: $t_{25} = 0.96$, $p = 0.34$, Fig. 4a, please also see Supplementary Table 8 for full results). The learning rates associated with negative feedback were not different across the conditions (Fig. 4b).

In general, a win–stay strategy is an adaptive one because, by definition, it leads to repetition of a previously successful choice. Such a strategy may not be adaptive, however, after reward is experienced for choosing an option with a low average rate of reward. In fact, because the high probability option is not presented on every trial, participants had to make choices between two options with a low average rate of reward on some trials. We, therefore, also investigated whether implementing a higher rate of win–stay strategy could also lead participants to engage in maladaptive win–stay behaviour. For example, consider the case where, first, on one trial, the participant chooses between two low probability options and this trial is then followed by one in which the low probability option is presented with a high probability option. In this case, even if 30% of the first trials lead to a positive outcome, the best strategy would be to ignore the positive outcome and to switch to the high probability option on the subsequent trial. We found a main effect of condition (ANOVA: $F_{2,75} = 3.75$, $p = 0.02$) and indeed found that participants did not repeat the same choice on the subsequent trial when a high probability symbol was on the screen in the Sham condition ($t_{25} = -0.3$, $p = 0.76$), and in the TUS-dACC condition ($t_{25} = 0.85$, $p = 0.39$), but this was not the case after TUS of the NAcc ($t_{25} = 3.24$, $p = 0.003$; Fig. 4c). In this situation, participants were more likely to stay after a win in a loss context (when choosing between two low probability options). Supplementary Table 9 presents all the results for this analysis.

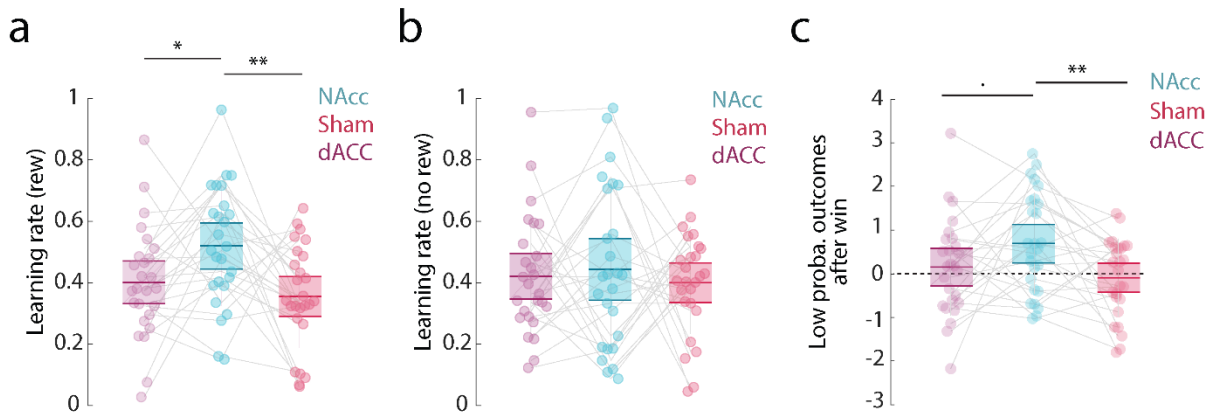

Fig. 4. Learning rates from RL model and maladaptive behaviours. a Higher learning rates after positive feedback in the TUS-NAcc condition. **b** No change in learning rates after negative feedback after TUS-NAcc. **c** Increase in rate of repetition of low probability options after reward showing maladaptive behaviour after TUS-NAcc condition compared to both TUS-dACC and Sham conditions.

FMRI findings

Our behavioural analyses revealed an overall increase in reward sensitivity which we explored in our main fMRI analysis by investigating the parametric BOLD responses to both reward expectation and reward delivery. We also constructed a parametric prediction error signal by contrasting reward delivery and reward expectation as per the RL formula (see Methods) without compromising the collinearity of our model. Our main hypothesis was that there would be a task-related change in BOLD in the region targeted with TUS, which we first investigated with an ROI analysis (see Methods and Supplementary Figure 1). This analysis revealed a clear increase in the parametric BOLD response to reward expectation in the NAcc in the TUS-NAcc condition compared to the Sham and TUS-dACC conditions (ANOVA: $F_{2,75} = 7.15$, $p = 0.001$; post hoc t-tests: dACC – NAcc: $t_{25} = -3.92$, $p = 0.0006$; Cohen D = 0.77; NAcc – sham: $t_{25} = -2.38$, $p = 0.024$; Cohen D = 0.47; dACC – sham: $t_{25} = -1.34$, $p = 0.19$; Cohen D = 0.26). This is displayed in Figure 5a and reported in full in Supplementary Table 10. A similar ROI analysis revealed a statistically significant variation in parametric BOLD activity linked to reward delivery in the dACC across conditions (ANOVA: $F_{2,75} = 3.12$, $p = 0.049$), however, the t-tests revealed that the comparison between TUS-dACC and sham was not significant ($t_{25} = -1.47$, $p = 0.15$; Fig. 5b), although TUS-dACC and TUS-NAcc was ($t_{25} = -2.90$, $p = 0.007$). It is important to note that the ANOVA shows a variation between conditions but does not identify which conditions drive the effect.

This ROI analysis was followed by a whole brain repeated measures ANOVA which allowed us to investigate the full spatial extent of the difference between the TUS conditions. This

analysis confirmed an increase in BOLD signalling of reward expectation after TUS-NAcc compared to Sham, not only in the NAcc, the sonicated region, but also in adjacent striatum, the thalamus, amygdala, precuneus, and posterior cingulate cortex ($Z > 2.3$; FDR corrected at $p < 0.05$; Fig. 5c, left panel). This analysis also revealed a statistically significant increase in the correlation between BOLD and reward expectation after TUS-NAcc compared to TUS-dACC not just in the NAcc but also in the vmPFC, lOFC, insula, thalamus, putamen, precuneus, and PCC ($Z > 2.3$; FDR-corrected at $p < 0.05$; Fig. 5c, middle panel). In the dACC, however, there was no statistically significant difference for either reward expectation or reward delivery between the TUS-dACC and Sham conditions (Fig. 5c, right panel).

However, a whole-brain repeated measures ANOVA did identify an increase in BOLD signalling of reward delivery after TUS-dACC compared to Sham in a distributed network adjacent to the region of sonication, in adjacent dACC and medial prefrontal cortex ($Z > 2.3$; FDR-corrected at 0.05; Fig. 5d, right panel). Thus, although the previous ROI analysis conducted precisely at the focus of the TUS had not revealed any difference between TUS-dACC and sham, the immediately adjacent dACC and medial prefrontal cortex did exhibit changes when the whole brain analysis was performed. The whole brain approach also revealed a stronger correlation between BOLD activity and reward delivery after TUS-dACC compared to TUS-NAcc in a distributed network, particularly surrounding the TUS focus, including the TUS focus itself, as shown in the previous ROI analysis ($Z > 2.3$; FDR-corrected at 0.05; Fig. 5d, middle panel; compare dACC and NAcc bars in 5b). There was no statistically significant difference between reward delivery-related activity after TUS-NAcc and Sham (Fig. 5d, left panel).

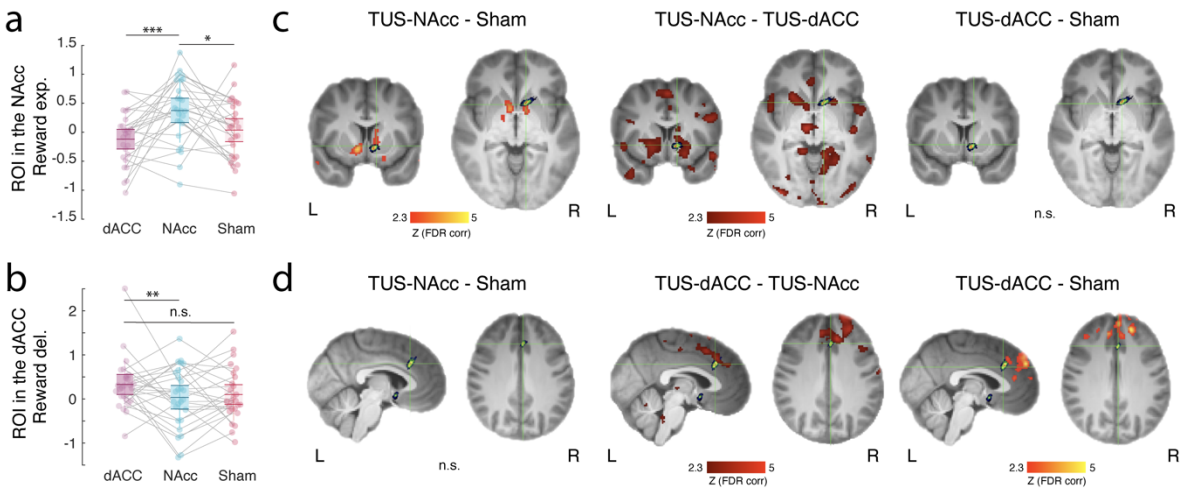

Fig. 5. ROI-based analysis and whole-brain differences. a Enhancement of the parametric BOLD representation of reward expectation in the NAcc after TUS-NAcc compared to Sham and TUS-dACC. **b** Although a main effect

was observed in the dACC ROI for an increase in the BOLD representation of reward delivery after TUS-dACC, this effect was not significantly different from that seen in the Sham condition in the post-hoc test. TUS-dACC compared to TUS-NAcc was, however, significant. **c** The same analysis at the whole brain level revealed that TUS-NAcc not only increased the NAcc signalling of reward anticipation in the ROI but also in adjacent striatum. Evidence of an even more distributed network of areas in which reward signalling was more prominent became apparent when TUS-NAcc was compared to TUS-dACC. However, consistent with panel **b**, reward expectation-related activity remained prominent in the NAcc and adjacent striatum. **d** Similarly, when contrasting the whole brain maps between TUS-dACC to TUS-NAcc and Sham, we found strong evidence for BOLD changes in dACC and medial frontal regions adjacent to the sonication site. This was not the case when contrasting TUS-NAcc to sham.

We then re-ran the same analysis but this time, rather than looking at the differences between the conditions, we focused on the main parametric effects of reward expectation and reward delivery within each of the three conditions separately. Regarding reward expectation (Fig. 6, left panel), the whole-brain analysis indicated a large, distributed network response to reward expectation in the vmPFC, lOFC, PCC, and insula in the Sham condition. This network was similar in nature, but activity was, in general, higher and more widely distributed after TUS-NAcc. We also found a parametric representation of reward expectation in prefrontal regions after TUS-dACC, although not as distributed as the other two maps. In terms of reward delivery (Fig. 6, right panel), whole-brain analysis confirmed previous findings and showed a distributed network response to reward delivery centred in the NAcc, in the Sham condition. This network was similar in nature after TUS-NAcc. However, we found that reward delivery was associated with an increased signal in many prefrontal and posterior regions after TUS-dACC.

Side effects associated with TUS

On the day of the stimulation, as well as on the following day, participants were prompted to report any adverse events resulting from the TUS–MRI intervention by filling out a 4-point rating scale questionnaire (see Methods). We found no difference between the TUS sessions and the Sham condition on the day of, and the day after, the session. Results are presented in Supplementary Figure 2. Upon qualitative inspection, it became evident that participants experienced various MRI-related issues, including sleepiness, neck pain, headache, and blurry vision, which were attributed to the prolonged period of lying down and the lighting conditions inside the MR scanner. However, there were also a few events that participants attributed to the TUS intervention. Specifically, one individual reported a mild headache around the region of stimulation after the TUS-dACC session, necessitating the use of paracetamol. Another participant, following the TUS-NAcc session, reported a very mild ache on the left side of the

forehead above the eyebrow which disappeared the following day. During Sham TUS, one person reported a slight blurriness of vision, which disappeared after the simulated sonication ended.

Discussion

This study aimed to capitalize on the high spatial resolution of TUS and its capacity to reach deep regions in the brain to target the NAcc in humans in the context of probabilistic reversal learning. A total of 26 healthy participants were enrolled in a within-subject repeated measures design experiment involving repetitive TUS and subsequent fMRI. After the application of 5Hz patterned TUS for 80 seconds in a counterbalanced fashion to the NAcc (TUS-NAcc), the dACC (TUS-dACC), or no sonication (Sham), participants performed a probabilistic reversal learning task in the MRI scanner which started on average 15min post sonication, when any potential auditory or somatosensory effects of stimulation were dissipated. We used both direct measures of reward sensitivity and model-based estimates of the expected value associated with each potential choice stimulus. The models were also used to examine prediction errors when participants received feedback to indicate if the choice was rewarded or unrewarded in analyses of both behaviour and neural activity.

With careful individualised TUS planning, using an estimate of each participant's skull image to achieve an optimised trajectory, it was possible to show that TUS-NAcc has neural effects that are most prominent in the region stimulated and that they are associated with changes in indices of behaviour that are similar to those emphasized in previous NAcc lesion studies [37–40]. Indeed, we found significant alterations in reward-related behaviours, including alterations in the tendency to adopt a win–stay strategy, changes in learning rates following rewards, and a changed learning curve for the rewarding option.

Unlike in lesion studies, however, where animals tended to become less sensitive to reward, TUS-NAcc had the opposite effect and led to an enhancement of each of the indices of the influence of reward on behaviour. The enhancement effect may reflect some feature of NAcc organization or, possibly, that the more conservative TUS parameters employed in the current study (in contrast to previous animal studies) led to an enhancement rather than a diminution in reward sensitivity. Exactly which TUS parameters are crucial for determining whether TUS exerts enhancing or disruptive effects remains to be determined. This will be particularly important for the potential use of TUS in patients and in disorders related to reward including those where reward sensitivity may be bidirectionally disrupted (e.g., substance use or bipolar

disorder), or those where reward sensitivity may be reduced (e.g., depression). Given the great interest in the possibility of TUS-based therapy for depression[246], careful consideration of whether TUS can impact reward processing is merited.

The one deficit in behaviour that was found after TUS-NAcc was, in fact, also due to increased reward sensitivity: participants exhibited a greater inclination to repeat a choice after receiving a reward that should have been ignored. We also found that TUS-NAcc induced neural changes in the parametric BOLD response associated with reward expectation in the NAcc and adjacent regions. Interestingly, BOLD changes linked to reward expectation in interconnected areas, even those at a distance, were also observed.

Lesion studies in macaques have similarly underlined the particular importance of the NAcc and adjacent ventral striatum in learning the values of states regardless of how those states are defined, although stimulus-reward learning rather than action-reward learning is most affected [37]. Although the ventral striatum seems to respond to both aversive stimuli as well as rewarding stimuli, lesions only affect reward-based learning [39]. Ventral striatal activity is present in deterministic and stochastic learning situations, but only the latter are affected by ventral striatal lesions [38]. More generally, the results align with prior studies suggesting that different families of neurons are linked to learning from reward and punishment [47].

The NAcc is a key component of the brain's reward circuitry and a main target for dopaminergic projections from the midbrain, particularly the ventral tegmental area and substantia nigra compacta. In healthy humans, this key reward-related brain region has been shown to signal reward expectation in BOLD changes and to possess activity correlated with changes in learning [48].

This study provides evidence that TUS can be used in humans to manipulate a deep subcortical region and test, through cognitive and computational models, its precise relationship with, and causal importance for, downstream behaviours. Many deep subcortical regions and their subdivisions play a crucial role in regulating fundamental behaviours and cognitive functions [49]. These regions are often difficult to access using traditional neuromodulation techniques and testing their causal roles in humans remains largely unexplored. Therefore, this study opens up a potentially new and large space in which to examine hypotheses about human brain activity and its relationship with cognition and behaviour, with important lessons for future studies in neuropsychiatric conditions. In this work, we also show that TUS to the NAcc did not cause general changes in performance and reaction time despite altering reward processing. This shows that TUS effects can be circumscribed to

the functional role of the region sonicated while global performance and general aspects of behaviour remain preserved.

To carry out this work, we planned the TUS intervention on each individual participant to maximise target engagement and minimise losses through the skull while ensuring safety prior to conducting the three TUS sessions. This was achieved by carefully positioning the TUS on each individual skull while assessing TUS trajectory and in-situ intensity and pressure. Although this was planned using a personalised pseudo-CT [19], it is possible that other methods may achieve higher precision, for example using an ultrashort echo time image (e.g., PETRA on Siemens) or computerized tomography (CT) scans. Using personalised TUS planning allowed us to carefully engage the NAcc and no other regions. Similarly, the resolution of TUS appears adequate for investigation of even smaller regions than the one investigated here, for example, such as teasing apart different roles of the NAcc core and shell [50]. Careful application of TUS to different regions in each hemisphere might also make it possible to test the causal role of a distributed circuit using the logic of disconnection studies [51,52].

Dysfunction in the reward system, including in the NAcc, is implicated in various psychiatric disorders such as addiction [53,54], depression [55], and certain eating disorders [56,57]. Methods for modulating the NAcc often involve invasive procedures, such as deep brain stimulation. In this study, we have shown that TUS can be both safe and effective in altering reward sensitivity non-invasively. The reduced risks associated with this approach suggest that it might be adapted for therapeutic interventions and clinical translation as well as providing a tool for researchers studying the neural underpinnings of these conditions. However, as with any new tool in the clinic, there will be a need for careful consideration of the ethics of TUS use and the establishment of safety guidelines.

One limitation of the approach used in the current study is that the same free field intensity was applied to all participants. In the future, however, it may be possible to develop individualized neuromodulation protocols to ensure that, instead, the in-situ intensity for each participant is constant. Furthermore, customization of TUS parameters based on an individual's baseline reward sensitivity might also significantly enhance efficacy. Such personalized approaches may be important not just in research but also in therapeutic settings.

Methods

Participants

Twenty-six healthy volunteers (14 female) aged between 20 and 65 years (mean = 36.3, s.d. ± 13.3) participated in the study. Participants were screened for contraindications to TUS and MRI [20], had no current diagnosis of neurological or psychiatric disorders, and were free of psychoactive medications at the time of the study. Written informed consent was obtained from all participants after experimental procedures were explained in full. The study was approved by the University of Plymouth Faculty of Health Staff Research Ethics and Integrity Committee (reference ID: 2487; date: 13/12/2021). Participants received £110 in total including a performance bonus of £10 for completing all study sessions. Travel expenses were also reimbursed up to £10 per session. All study sessions took place at the Brain Research & Imaging Centre in Plymouth, United Kingdom.

The study design is summarised in Fig. 2a. Participants completed a behavioural practice and MRI only session followed by three TUS and MRI sessions, which were spaced at least one week apart and at the same time of day for each participant. During the behavioural practice and MRI session, participants familiarised themselves with the reversal learning task and underwent a short series of MRI scans including a high-resolution T1-weighted MRI. This anatomical scan was used to derive a participant-specific head model for neuronavigation and acoustic simulations to plan TUS target and transducer placement for the subsequent TUS and MRI sessions. Participants were then assigned a randomised order of TUS conditions for their TUS and MRI sessions. During the TUS and MRI sessions, participants underwent 80 s of TUS followed by a series of MRI scans including interleaved blocks of fMRI and MR spectroscopy scans during which they performed the task (Fig. 2b). MR spectroscopy data will be the subject of another report. The TUS conditions were either active TUS applied to the left nucleus accumbens (TUS-NAcc), active TUS applied to the left dorsal anterior cingulate cortex (TUS-dACC), or Sham, where the transducer was placed as if to target the medial frontal cortex, but no ultrasound was delivered (Fig. 2c).

TUS protocol and procedure

Target location for ultrasound

The left NAcc target was centred at Montreal Neurological Institute (MNI) coordinates x = -9, y = 11, z = -7 and was identified based on an 80% probability atlas from the Harvard–Oxford

Subcortical Structural Atlas supplied with FSL. The left dACC target was centred at MRI coordinates x = -5, y = 24, z = 30. During target and transducer placement planning, the targets were identified using an initial co-registration to MNI space and adjusted based on each individual's T1-weighted MRI.

Acoustic simulations

The output of our NeuroFUS transducer was previously measured with a hydrophone setup in a water tank (see Yaakub and colleagues [20] for details). We used the k-Wave Toolbox [58] (version 1.4) and custom scripts [19,20] in MATLAB (R2020b, MathWorks, Inc.) for our simulations, with an individual skull model estimated from each participant's T1-weighted MRI [19,20]. The simulated transducer was modelled based on the physical properties of the NeuroFUS CTX-500-4 transducer and optimised phases obtained using k-Plan (https://dispatch.k-plan.io). We set our simulation grid size to a 256 × 256 × 256 matrix centred on the midpoint between the transducer and focus with a grid spacing of 0.5 mm (i.e., 6 points per wavelength at 500 kHz).

TUS protocol

The NeuroFUS TPO system (Brainbox Ltd., Cardiff, UK) with a four-element annular transducer (CTX-500-4, diameter = 64 mm, central frequency = 500 kHz) was used to deliver 5 Hz pulse repetition frequency repetitive TUS (pulse duration = 20 ms, pulse repetition interval = 200 ms, total duration = 80 s, total number of pulses = 400). The target free field spatial-peak pulse-average intensity (I_{SPPA}) was kept constant at 30 W/cm^2 for all participants, which is the intensity before going through the skull. Transcranial acoustic and thermal simulations (see "Acoustic simulations" section for details) were performed both during planning and after each session to confirm that transcranial intensities remained within the limits of the FDA safety guidelines for diagnostic ultrasound (MI ≤ 1.9; I_{SPPA} ≤ 190 W/cm^2) and the maximum temperature rise in the brain did not exceed 2°C anywhere in soft tissue.

To ensure good ultrasound transmission, we spread a layer of ultrasound transmission gel (Aquasonic 100, Parker Laboratories Inc.) into the hair where the transducer would be placed, making sure that air bubbles were smoothed or combed out. The head was not shaved. A 2 cm gel pad (Aquaflex, Parker Laboratories Inc.) was used between the transducer and the head. Neuronavigation was performed with Brainsight v 2.5 (Rogue Research Inc., Montréal, Québec, Canada) using the anatomical T1-weighted MR images from each participant. The

focal depth read off Brainsight during each session was entered on the NeuroFUS TPO before stimulation was delivered. Once stimulation had begun, the trajectory was sampled with Brainsight and used in confirmatory acoustic simulations performed after each session. At the end of and on the day after each TUS session, participants were asked to report and elaborate on any adverse effects they thought were associated with TUS using a 4-point scale questionnaire with open ended responses where they were encouraged to further describe any experiences and whether they thought they experiences were related to the study procedures (provided in the Supplementary Material).

Sham TUS was delivered in the same way as active TUS, except that the NeuroFUS TPO unit was turned off and no stimulation was delivered. A sound mimicking the sound produced by the transducer was played via bone conduction headphones (see [20] for details) placed bilaterally approximately 2 cm posterior and superior to the temples. They were worn for all the conditions, including the NAcc and dACC active TUS but the sound was not played during those. We verbally prompted participants at the end of the study to disclose whether they had felt any difference between the sessions. We found that participants were unable to discern any differences between the sessions and had not suspected a sham condition.

MRI data acquisition and pre-processing

MRI scans were acquired on a Siemens MAGNETOM Prisma 3T scanner (VE11E, Siemens Healthineers, Erlangen, Germany) with a 32-channel head coil. The scans in this study included a T1-weighted MPRAGE sequence acquired in the sagittal plane (repetition time (TR) = 2100 ms, echo time (TE) = 2.26 ms, inversion time = 900 ms, flip angle (FA) = 8°, GRAPPA acceleration factor = 2, field of view = 256 × 256 mm, number of slices = 176, voxel size = 1 × 1 × 1 mm^3), two GE-EPI fMRI scans during which participants performed the probabilistic reversal learning task lasting approximately 10-minutes each (acquisition plane tilted 30° clockwise from the line parallel to the AC-PC line, 1400 ms TR, 30 ms TE, 67° FA, 2.5 mm slice thickness, no slice gap, multi-band acceleration factor of 2, and 60 interleaved slices of 96 × 96 matrix size, giving a voxel size of 2.5 × 2.5 × 2.5 mm^3), and field maps for fMRI distortion correction. fMRI pre-processing was performed using tools from the FMRIB Software Library v6.0 (FSL; www.fmrib.ox.ac.uk/fsl). Pre-processing included MCFLIRT motion correction, B0 inhomogeneity correction (Effective EPI echo spacing = 0.49 ms, EPI TE = 30 ms, unwarp dif = -y, signal loss thr = 10%), brain extraction, spatial smoothing (5 mm FWHM), and high pass filtering (0.01 Hz).

Probabilistic reversal learning task

Participants performed a probabilistic reversal learning task during four blocks of MR acquisitions (two of which were fMRI acquisitions). The task consisted of two runs of 100 trials each (presented during the fMRI scans) and two runs of 60 trials each (no fMRI), giving a total of 320 trials across four blocks. A set of 12 abstract symbols were used in the experiment (adapted from [31]). Additional stimuli included a tick to represent a reward, a cross to represent no reward, and a fixation cross.

In each task block, a different subset of three abstract symbols (e.g. A, B and C) was randomly selected from the full set of 12 symbols. One of the three symbols would be associated with a 70% chance of obtaining a reward ("high" reward probability symbol) while the remaining two symbols were each associated with a 30% chance of obtaining a reward ("low" reward probability symbols). During each trial, participants were shown two of the three symbols and asked to select the symbol that they thought was associated with the highest probability of obtaining a reward. Participants were not informed of the exact reward probabilities assigned to each symbol but were instead asked to learn to choose the symbol that was more likely to lead to a reward through trial and error (i.e., by making use of the outcome of past decisions). Therefore, there was only one option associated with a high reward rate, while the other two were associated with lower reward rates. This procedure has a few advantages. First, because the best option is not presented on every trial, participants had to make different type of choices on different trials depending on which options were available. In addition, because for a third of the trials, participants had to choose between the two least rewarding symbols, even after they have learnt the cue-reward contingencies, this manipulation allowed us to achieve a more even distribution of positive and negative outcomes.

After making their decision, participants were shown either a tick to represent a reward on that trail or a cross to represent no reward. Participants were told that the number of rewarded trials would be counted across all the blocks and sessions, and that they would be awarded up to £10 as a performance bonus at the end of the study depending how well they performed during the task. Participants were not informed of the exact reward probabilities assigned to each symbol but were instead asked to learn to choose the symbol that was more likely to lead to a reward through trial and error (i.e., by making use of the outcome of past decisions). To prevent participants from searching for non-existent patterns and to reduce cognitive load we presented the three possible pair combinations of the three symbols in a fixed order (i.e., AB, BC and CA). However, the presentation of the symbols on the left or right of the fixation cross

each time was randomised. Participants were explicitly informed about this manipulation. After every 25 trials in each block, a “reversal” would be introduced whereby the high reward probability was reassigned to a different symbol out of the three for that block. Participants would then have to learn to identify the “new” high probability symbol out of the three. Participants were informed that reversals would happen several times during each block but were not told of the exact frequency of reversals. The stochastic nature of the task and the reversals in choice-reward associated also ensured a need for constant assessment of each option’s value and constant changes in choice selection.

Fig. 2b shows the sequence of events for an example trial. At the start of each block of trials, the three abstract symbols selected for that block were shown on the screen for 5 s so that participants could familiarise themselves with the symbols they would see for the duration of that block. Each trial began with a fixation cross shown for a random delay of 1–1.2 s. This was followed by the presentation of two of the three symbols on either side of the fixation cross for 1.25 s, during which time participants were instructed to select one of the symbols by pressing a button. The fixation cross flickered for 100 ms after participants made their selection to indicate that their response was registered. The fixation cross was then shown again with a random delay of 1–1.2 s before the outcome of the trial, either a tick for a reward or a cross for no reward, was shown in the centre of the screen for 0.75 s. Trials in which participants failed to respond within 1.25 s were followed by a “Lost trial” message shown in the middle of the screen in place of the tick or cross outcome.

Stimuli display

The probabilistic reversal learning task was presented on an MRI-compatible LCD screen (BOLDscreen 32 AVI, 32-inch screen, resolution = 1920 × 1080, refresh rate = 120 Hz, Cambridge Research Systems Ltd., Rochester, Kent, UK) placed one meter behind the MRI scanner, which participants viewed using a mirror attached to the head coil. The experiment was presented using the Presentation software (Neurobehavioural Systems Inc., Berkeley, CA, USA) run on a Windows machine. Responses were collected using an MRI-compatible fibre optic response pad (model: HHSC-2x4-C, Current Designs Inc., Philadelphia, PA, USA) placed in the participants’ right hand. Participants selected the symbol on the right with their index finger and the symbol on the left with their middle finger.

Modelling of behavioural data

Models

Model 1 – To create trial-wise estimates of expected value (reward prediction) and prediction error, we first used a Rescorla–Wagner model, within the reinforcement learning (RL) framework, using each participants' behavioural choices and feedback. Specifically, the algorithm assigned each choice *i* (for example selecting the symbol A) an expected value $V_A(t)$ which was updated via a prediction error, $\delta(t)$, as follows:

$$V_A(t+1) = V_A(t) + \alpha * \delta(t),$$

where α is a learning rate that determines the influence of the prediction error on the updating of the symbol's expected value. The prediction error is calculated as:

$$\delta(t) = r(t) - V_A(t),$$

where $r(t)$ represents the outcome obtained on that trial (0 or 1). The expected values of the unselected stimulus (e.g., B) and the stimulus not shown on trial *t* (e.g., C) were not updated.

Model 2 - The main limitation of the classical Rescorla–Wagner model is the implied symmetry with which positive and negative feedback update a choice value estimate. This contradicts evidence that learning from positive and negative feedback has different effects on behaviour and decision-making [59]. We thus implemented an alternative asymmetric learning model which discriminates based on outcome valence, with two learning rates, one for positive and one for negative feedback [34,60].

With all models we used a SoftMax decision function in which, on each trial *t*, a stimulus choice probability (e.g., A) was given by:

$$P_A(i) = \sigma(\beta(V_A(t) - V_B(t)))$$

where $\sigma(z) = 1/(1 + e^{-z})$ is the logistic function, φ denotes the indecision point (when choosing each of the alternatives is equiprobable), and β represents the degree of choice stochasticity (i.e., the exploration/exploitation parameter). Choice probability of the unchosen stimulus (e.g., B) and of the stimulus not shown on trial t (e.g., C) were not updated.

Model fitting and comparison of models

We estimated parameters individually for each participant, block and stimulation condition (TUS-NAcc, TUS-dACC, and Sham) including β and the single or multiple learning rate α. We

initially determined reasonably good parameters by a grid search while applying the following parameter constraints: $\beta > 0$ and $0.1 < \alpha < 0.9$. The best parameters from the grid search were then used as starting points for a simplex optimization procedure, which determined the final parameter estimates. As a goodness-of-fit measure, we used the log likelihood of the observed choices over all trials *T* given the model and its parameters: $LL = \sum_{t=1}^{T} ln\,[ft(y/\theta)]$, where $ft(y/\theta)$ denotes the probability of choice *y* in trial *t* given the model's parameter set *θ*. Predicted choice probabilities were calculated based on 1000 simulations per parameter set (combinations of the free parameters), whereby in each simulation the model determined the choices used to update reward expectations (as opposed to observed choices). To obtain estimates of expected values and prediction errors, the two estimated participant-specific parameters were then re-entered into the reinforcement learning algorithm, this time based on participants' observed choices.

To determine the best fitting model, we performed classical model comparison. Specifically, for each model we first estimated the subject-wise Bayesian Information Criterion (BIC) as follows:

$$BIC = -2log\,L + dlogn$$

Here, the goodness of fit of a model ($-log\,L$) is penalised by the complexity term ($dlogn$) where the number of free parameters in the model *d* is scaled by the number of data points *n* (i.e., trials). We then computed the sum of the subject-wise BIC for each model and compared the model-wise BIC estimates (lower estimates indicating better fit).

Behavioural analysis

Behavioural data were saved in standard ascii data files and were analysed using custom-written code in MATLAB (R2023a, MathWorks, Inc., https://uk.mathworks.com/). The relationships between Reward, Prediction Error, win–stay behaviours, and task structure were evaluated using linear regression. Statistical testing and post hoc analyses were performed using the default functions in MATLAB, namely *anova, fitglm, fitlme* and *ttest*. We ran two series of analyses, very similar in nature. The first one was run without estimates from the RL model, simply the behaviours during the task. The second series used estimates from the RL model.

Analysis A

The first analysis in the series was concerned with the relationship between reward and subsequent win–stay behaviours. We thus ran a regression model given by the equation:

$$WinStay\ (t+1) \sim Reward\ (t)\ +\ choiceStickiness\ (t)\ +\ isHPScreen\ (t+1)$$

where $WinStay\ (t+1)$ is coded as +1 for Win–Stay, -1 for Win–Switch and 0 otherwise on trial t+1; Reward is coded as 1 if a reward is received and 0 if not on trial t; *choiceStickiness* is coded as 1 if choice has been repeated and 0 otherwise on trial t; and *isHPscreen* is coded as 1 if there was a High Probability Symbol on the screen and 0 otherwise. In essence, this is equivalent to looking at the proportion of win–stay behaviours while controlling for other features of the task and behaviour such as any general tendency to choose repetition regardless of reward (*choiceStickiness*). We then subjected the regression weights to an analysis of variance (ANOVA) such that:

$$regressionCoef \sim 1\ +\ condition$$

where condition is a categorical variable (TUS-dACC, TUS-NAcc, or Sham). Further three two-tailed two-sample t-tests were used for all possible TUS-dACC, TUS-NAcc and Sham pairs applying Bonferroni correction. Cohen D was calculated for a paired-samples t-test by dividing the mean difference by the standard deviation of the difference.

Analysis B

The second analysis tested the difference between the regression weights from Analysis A for NAcc vs. Sham and dACC vs. Sham for each of the four blocks (2 fMRI and 2 tasks only). As these blocks were acquired at different times, this allowed us to plot the effect identified in Analysis A over time. We used two-tailed paired t-tests to check significance of this difference for each block and considered $p < 0.05$ to be significant. We did not perform multiple comparisons as we used predefined hypotheses and tests.

Analysis C

The third analysis looked at the rate of choice towards the high probability symbol, for the window significant in Analysis B, for the runs within a block, averaged across blocks, and then during the first run of a block, averaged across blocks.

We then repeated the Analyses A, B, and C with model estimates from the RL model. In Analysis A, we used the prediction error at t instead of the reward at t. In Analysis C, we used

the expected value associated with the high probability symbol instead of the choice rate associated with the high probability symbol and presented it for all four blocks.

Task-based fMRI analysis

FMRI data were pre-processed and analysed using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Pre-processing included motion correction, B0 field inhomogeneity correction, brain extraction, spatial smoothing (5 mm FWHM) and high pass filtering (0.01 Hz). FMRI data were co-registered to the MNI standard space via a linear transform to the individual's high-resolution T1-weighted MRI and a non-linear transform to the MNI template.

ROI analyses

We hypothesised that we would find reward-related BOLD changes within the targeted areas of neuromodulation. For the NAcc site, we used the bilateral NAcc defined anatomically by the probabilistic Harvard–Oxford subcortical structural atlases. For the dACC site, we created a sphere around the maximum peak TUS intensity (ISPPA *in situ*) across participants extracted from our acoustic simulation (see Supplementary Figure 1).

Statistical analyses of BOLD data were then performed using a fixed-effects approach within the framework of a GLM, as implemented in FSL (using the FEAT module). For each block, we ran a GLM in the form:

$$BOLD \sim UnmodFeedback + RewDel + RewExp + ReactionTime + Lost + counfounds$$

where *BOLD* is a T × 1 (T time samples) column vector containing the times series data for a given voxel; *UnmodFeedback* is an unmodulated regressor (all event amplitudes set to 1) locked at the time of outcome (that is, when the tick/cross appeared) as a boxcar regressor with a duration of 100 ms; *RewDel* is a simple categorical regressor for reward delivery (amplitudes set to +1 for rewarded outcomes) as a boxcar regressor with a duration of 100 ms; *RewExp* is a fully parametric regressor whose event amplitudes were modulated by the expected probabilistic reward associated with the chosen option, locked at time of decision, as a boxcar regressor with a duration of 100 ms; *ReactionTime* is a boxcar regressor with a duration modulated by reaction time, locked at time of decision; and *Lost* is an unmodulated regressor for all lost trials, locked at time of decision. In addition, we included six nuisance regressors, one for each of the motion parameters (three rotations and three translations).

From each of the two reward-related regressors (*RewExp* and *RewDel*), we extracted beta coefficients from the NAcc and dACC by back-projecting the ROIs described above (Supplementary Figure 1) from standard space into each individual's EPI (functional) space by applying the inverse transformations as estimated during registration. For the two reward-related regressors, we computed the average beta coefficients from all voxels in the back-projected ROIs and across participants to test the overall BOLD response profile of the ROIs as a function of both reward delivery and reward expectation.

Whole-brain analyses

We ran a single-group tripled t-test which corresponds to a repeated measures ANOVA with one fixed factor with three levels and one random factor. Fitting such a mixed effects model with ordinary least squares (OLS) (as implemented in FEAT) requires an assumption of *compound symmetry*. This is the state of equal variance and intra-subject correlations being equal. That is, Cov(scan1,scan2) = Cov(scan1,scan3) = Cov(scan2,scan3). For these whole-brain fMRI results, all images were thresholded given a one-sided t-test and subsequently FDR-corrected at $p < 0.05$.

References

1. Silberberg, D., Anand, N.P., Michels, K., and Kalaria, R.N. (2015). Brain and other nervous system disorders across the lifespan - global challenges and opportunities. Nature *527*, S151-154. 10.1038/nature16028.

2. Mahoney, J.J., Thompson-Lake, D.G.Y., Ranjan, M., Marton, J.L., Carpenter, J.S., Zheng, W., Berry, J.H., Farmer, D.L., D'Haese, P., Finomore, V.S., et al. (2023). Low-Intensity Focused Ultrasound Targeting the Bilateral Nucleus Accumbens as a Potential Treatment for Substance Use Disorder: A First-in-Human Report. Biological Psychiatry *94*, e41–e43. 10.1016/j.biopsych.2023.06.031.

3. Yan, H., Shlobin, N.A., Jung, Y., Zhang, K.K., Warsi, N., Kulkarni, A.V., and Ibrahim, G.M. (2023). Nucleus accumbens: a systematic review of neural circuitry and clinical studies in healthy and pathological states. J Neurosurg *138*, 337–346. 10.3171/2022.5.JNS212548.

4. Mayberg, H.S., Lozano, A.M., Voon, V., McNeely, H.E., Seminowicz, D., Hamani, C., Schwalb, J.M., and Kennedy, S.H. (2005). Deep brain stimulation for treatment-resistant depression. Neuron *45*, 651–660. 10.1016/j.neuron.2005.02.014.

5. Yoo, S., Mittelstein, D.R., Hurt, R.C., Lacroix, J., and Shapiro, M.G. (2022). Focused ultrasound excites cortical neurons via mechanosensitive calcium accumulation and ion channel amplification. Nat Commun *13*, 493. 10.1038/s41467-022-28040-1.

6. Darmani, G., Bergmann, T.O., Pauly, K.B., Caskey, C.F., de Lecea, L., Fomenko, A., Fouragnan, E., Legon, W., Murphy, K.R., and Nandi, T. (2021). Non-invasive transcranial ultrasound stimulation for neuromodulation. Clinical Neurophysiology.

7. Murphy, K.R., Farrell, J.S., Gomez, J.L., Stedman, Q.G., Li, N., Leung, S.A., Good, C.H., Qiu, Z., Firouzi, K., Butts Pauly, K., et al. (2022). A tool for monitoring cell type–specific focused ultrasound neuromodulation and control of chronic epilepsy. Proceedings of the National Academy of Sciences *119*, e2206828119. 10.1073/pnas.2206828119.

8. Blackmore, D.G., Razansky, D., and Götz, J. (2023). Ultrasound as a versatile tool for short- and long-term improvement and monitoring of brain function. Neuron *111*, 1174–1190. 10.1016/j.neuron.2023.02.018.

9. Bault, N., Yaakub, S.N., and Fouragnan, E. (2024). Early-phase neuroplasticity induced by offline transcranial ultrasound stimulation in primates. Current Opinion in Behavioral Sciences *56*, 101370. 10.1016/j.cobeha.2024.101370.

10. Folloni, D., Verhagen, L., Mars, R.B., Fouragnan, E., Constans, C., Aubry, J.-F., Rushworth, M.F.S., and Sallet, J. (2019). Manipulation of Subcortical and Deep Cortical Activity in the Primate Brain Using Transcranial Focused Ultrasound Stimulation. Neuron *101*, 1109-1116.e5. 10.1016/j.neuron.2019.01.019.

11. Khalighinejad, N., Bongioanni, A., Verhagen, L., Folloni, D., Attali, D., Aubry, J.-F., Sallet, J., and Rushworth, M.F.S. (2020). A Basal Forebrain-Cingulate Circuit in Macaques Decides It Is Time to Act. Neuron *105*, 370-384.e8. 10.1016/j.neuron.2019.10.030.

12. Verhagen, L., Gallea, C., Folloni, D., Constans, C., Jensen, D.E., Ahnine, H., Roumazeilles, L., Santin, M., Ahmed, B., Lehericy, S., et al. (2019). Offline impact of transcranial focused ultrasound on cortical activation in primates. eLife *8*, e40541. 10.7554/eLife.40541.

13. Folloni, D., Fouragnan, E., Wittmann, M.K., Roumazeilles, L., Tankelevitch, L., Verhagen, L., Attali, D., Aubry, J.-F., Sallet, J., and Rushworth, M.F.S. (2021). Ultrasound modulation of macaque prefrontal cortex selectively alters credit assignment–related activity and behavior. Science Advances *7*, eabg7700. 10.1126/sciadv.abg7700.

14. Fouragnan, E.F., Chau, B.K.H., Folloni, D., Kolling, N., Verhagen, L., Klein-Flügge, M., Tankelevitch, L., Papageorgiou, G.K., Aubry, J.-F., Sallet, J., et al. (2019). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. Nature Neuroscience. 10.1038/s41593-019-0375-6.

15. Bongioanni, A., Folloni, D., Verhagen, L., Sallet, J., Klein-Flügge, M.C., and Rushworth, M.F.S. (2021). Activation and disruption of a neural mechanism for novel choice in monkeys. Nature *591*, 270–274. 10.1038/s41586-020-03115-5.

16. Kubanek, J., Brown, J., Ye, P., Pauly, K.B., Moore, T., and Newsome, W. (2020). Remote, brain region-specific control of choice behavior with ultrasonic waves. Sci Adv *6*, eaaz4193. 10.1126/sciadv.aaz4193.

17. Deffieux, T., Younan, Y., Wattiez, N., Tanter, M., Pouget, P., and Aubry, J.-F. (2013). Low-intensity focused ultrasound modulates monkey visuomotor behavior. Curr. Biol. *23*, 2430–2433. 10.1016/j.cub.2013.10.029.

18. Attali, D., Tiennot, T., Schafer, M., Fouragnan, E., Sallet, J., Caskey, C.F., Chen, R., Darmani, G., Bubrick, E.J., Butler, C., et al. (2023). Three-layer model with absorption for conservative estimation of the maximum acoustic transmission coefficient through the human skull for transcranial ultrasound stimulation. Brain Stimulation *16*, 48–55. 10.1016/j.brs.2022.12.005.

19. Yaakub, S.N., White, T.A., Kerfoot, E., Verhagen, L., Hammers, A., and Fouragnan, E.F. (2023). Pseudo-CTs from T1-weighted MRI for planning of low-intensity transcranial focused ultrasound neuromodulation: An open-source tool. Brain Stimulation *16*, 75–78. 10.1016/j.brs.2023.01.838.

20. Yaakub, S.N., White, T.A., Roberts, J., Martin, E., Verhagen, L., Stagg, C.J., Hall, S., and Fouragnan, E.F. (2023). Transcranial focused ultrasound-mediated neurochemical and functional connectivity changes in deep cortical regions in humans. Nature Communications *14*, 5318. 10.1038/s41467-023-40998-0.

21. Averbeck, B., and O'Doherty, J.P. (2022). Reinforcement-learning in fronto-striatal circuits. Neuropsychopharmacology *47*, 147–162. 10.1038/s41386-021-01108-0.

22. Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron *66*, 585–595. 10.1016/j.neuron.2010.04.016.

23. Liljeholm, M., and O'Doherty, J.P. (2012). Contributions of the striatum to learning, motivation, and performance: an associative account. Trends Cogn Sci *16*, 467–475. 10.1016/j.tics.2012.07.007.

24. Tobler, P.N., O'Doherty, J.P., Dolan, R.J., and Schultz, W. (2007). Reward Value Coding Distinct From Risk Attitude-Related Uncertainty Coding in Human Reward Systems. J Neurophysiol *97*, 1621–1632. 10.1152/jn.00745.2006.

25. Volkow, N.D., Wise, R.A., and Baler, R. (2017). The dopamine motive system: implications for drug and food addiction. Nat Rev Neurosci *18*, 741–752. 10.1038/nrn.2017.130.

26. Averbeck, B.B., and Costa, V.D. (2017). Motivational neural circuits underlying reinforcement learning. Nat Neurosci *20*, 505–512. 10.1038/nn.4506.

27. Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. Nature reviews neuroscience *17*, 183–195.

28. Sutton, R.S., and Barto, A.G. (1998). Introduction to reinforcement learning (MIT press Cambridge).

29. Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599. 10.1126/science.275.5306.1593.

30. D'Ardenne, K., McClure, S.M., Nystrom, L.E., and Cohen, J.D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science *319*, 1264–1267. 10.1126/science.1150605.

31. Fouragnan, E., Retzler, C., Mullinger, K., and Philiastides, M.G. (2015). Two spatiotemporally distinct value systems shape reward-based learning in the human brain. Nature Communications *6*, 8107. 10.1038/ncomms9107.

32. Fouragnan, E., Queirazza, F., Retzler, C., Mullinger, K.J., and Philiastides, M.G. (2017). Spatiotemporal neural characterization of prediction error valence and surprise during reward learning in humans. Scientific reports *7*, 1–18.

33. Fouragnan, E., Retzler, C., and Philiastides, M.G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. Human brain mapping *39*, 2887–2906.

34. Niv, Y., Edlund, J.A., Dayan, P., and O'Doherty, J.P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. J. Neurosci. *32*, 551–562. 10.1523/JNEUROSCI.5498-10.2012.

35. Palminteri, S., Khamassi, M., Joffily, M., and Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. Nat Commun *6*, 8096. 10.1038/ncomms9096.

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

37. Rothenhoefer, K.M., Costa, V.D., Bartolo, R., Vicario-Feliciano, R., Murray, E.A., and Averbeck, B.B. (2017). Effects of Ventral Striatum Lesions on Stimulus-Based versus

Action-Based Reinforcement Learning. J Neurosci *37*, 6902–6914. 10.1523/JNEUROSCI.0631-17.2017.

38. Costa, V.D., Dal Monte, O., Lucas, D.R., Murray, E.A., and Averbeck, B.B. (2016). Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning. Neuron *92*, 505–517. 10.1016/j.neuron.2016.09.025.

39. Taswell, C.A., Costa, V.D., Murray, E.A., and Averbeck, B.B. (2018). Ventral striatum's role in learning from gains and losses. Proc Natl Acad Sci U S A *115*, E12398–E12406. 10.1073/pnas.1809833115.

40. Taswell, C.A., Janssen, M., Murray, E.A., and Averbeck, B.B. (2023). The motivational role of the ventral striatum and amygdala in learning from gains and losses. Behav Neurosci *137*, 268–280. 10.1037/bne0000558.

41. Giarrocco, F., Costa, V.D., Basile, B.M., Pujara, M.S., Murray, E.A., and Averbeck, B.B. (2024). Motor System-Dependent Effects of Amygdala and Ventral Striatum Lesions on Explore-Exploit Behaviors. J Neurosci *44*, e1206232023. 10.1523/JNEUROSCI.1206-23.2023.

42. Vicario-Feliciano, R., Murray, E.A., and Averbeck, B.B. (2017). Ventral striatum lesions do not affect reinforcement learning with deterministic outcomes on slow time scales. Behav Neurosci *131*, 385–391. 10.1037/bne0000211.

43. Klein-Flügge, M.C., Fouragnan, E.F., and Martin, E. (2024). The importance of acoustic output measurement and monitoring for the replicability of transcranial ultrasonic stimulation studies. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation *17*, 32–34. 10.1016/j.brs.2023.12.002.

44. Aubry, J.-F., Attali, D., Schafer, M., Fouragnan, E., Caskey, C., Chen, R., Darmani, G., Bubrick, E.J., Sallet, J., Butler, C., et al. (2023). ITRUSST Consensus on Biophysical Safety for Transcranial Ultrasonic Stimulation. Preprint at arXiv, 10.48550/arXiv.2311.05359 10.48550/arXiv.2311.05359.

45. Fouragnan, E., Retzler, C., Mullinger, K., and Philiastides, M.G. (2015). Two spatiotemporally distinct value systems shape reward-based learning in the human brain. Nature Communications *6*, 8107. 10.1038/ncomms9107.

46. Philip, N.S., and Arulpragasam, A.R. (2022). Reaching for the unreachable: low intensity focused ultrasound for non-invasive deep brain stimulation. Neuropsychopharmacology, 1–2. 10.1038/s41386-022-01386-2.

47. Matsumoto, M., and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature *459*, 837–841. 10.1038/nature08028.

48. Haber, S.N., and Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human Imaging. Neuropsychopharmacol *35*, 4–26. 10.1038/npp.2009.129.

49. Klein-Flügge, M.C., Jensen, D.E., Takagi, Y., Priestley, L., Verhagen, L., Smith, S.M., and Rushworth, M.F. (2022). Relationship between nuclei-specific amygdala connectivity and mental health dimensions in humans. Nature human behaviour *6*, 1705–1722.

50. Tian, Y., Margulies, D.S., Breakspear, M., and Zalesky, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. Nat Neurosci *23*, 1421–1432. 10.1038/s41593-020-00711-6.

51. Pujara, M.S., Ciesinski, N.K., Reyelts, J.F., Rhodes, S.E.V., and Murray, E.A. (2022). Selective Prefrontal-Amygdala Circuit Interactions Underlie Social and Nonsocial Valuation in Rhesus Macaques. J Neurosci *42*, 5593–5604. 10.1523/JNEUROSCI.0794-21.2022.

52. Wilson, C.R.E., Gaffan, D., Browning, P.G.F., and Baxter, M.G. (2010). Functional localization within the prefrontal cortex: missing the forest for the trees? Trends Neurosci *33*, 533–540. 10.1016/j.tins.2010.08.001.

53. Komarnyckyj, M., Retzler, C., Cao, Z., Ganis, G., Murphy, A., Whelan, R., and Fouragnan, E.F. (2022). At-risk alcohol users have disrupted valence discrimination during reward anticipation. Addict Biol *27*, e13174. 10.1111/adb.13174.

54. Komarnyckyj, M., Retzler, C., Whelan, R., Young, O., Fouragnan, E., and Murphy, A. (2023). Anticipatory reward dysfunction in alcohol dependence: An electroencephalography monetary incentive delay task study. Addiction Neuroscience *8*, 100116. 10.1016/j.addicn.2023.100116.

55. Queirazza, F., Fouragnan, E., Steele, J.D., Cavanagh, J., and Philiastides, M.G. (2019). Neural correlates of weighted reward prediction error during reinforcement learning classify response to cognitive behavioral therapy in depression. Sci Adv *5*, eaav4962. 10.1126/sciadv.aav4962.

56. Neveu, R., Neveu, D., Barsumian, F., Fouragnan, E., Carrier, E., Lai, M., Sultan, J., Nicolas, A., and Coricelli, G. (2014). Improved Planning Abilities in Binge Eating. PLOS ONE *9*, e105657. 10.1371/journal.pone.0105657.

57. Neveu, R., Fouragnan, E., Barsumian, F., Carrier, E., Lai, M., Nicolas, A., Neveu, D., and Coricelli, G. (2016). Preference for Safe Over Risky Options in Binge Eating. Front. Behav. Neurosci. *10*. 10.3389/fnbeh.2016.00065.

58. Treeby, B.E., and Cox, B.T. (2010). k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. Journal of Biomedical Optics *15*, 021314. 10.1117/1.3360308.

59. van den Bos, W., Güroğlu, B., van den Bulk, B.G., Rombouts, S.A.R.B., and Crone, E.A. (2009). Better than Expected or as Bad as You Thought? The Neurocognitive Development of Probabilistic Feedback Processing. Front Hum Neurosci *3*, 52. 10.3389/neuro.09.052.2009.

60. Gershman, S.J. (2015). Do learning rates adapt to the distribution of rewards? Psychon Bull Rev *22*, 1320–1327. 10.3758/s13423-014-0790-3.

Data and Software Availability

FSL can be downloaded from https://fsl.fmrib.ox.ac.uk/fsl/fslwiki. The data reported in this paper, including the behavioural and neural data is available on the OSF open platform (https://osf.io/j34qz/, https://osf.io/w3mev/, and https://osf.io/vst9y/) and as reported in Supplementary Material 1 (Resource Table). The code to generate pseudo-CTs from T1-weighted images and the in-house acoustic simulation code based on k-Wave (Matlab) can both be found on GitHub (https://github.com/sitiny/mr-to-pct and https://github.com/sitiny/BRIC_TUS_Simulation_Tools, respectively). K-wave can be downloaded from http://www.k-wave.org/. The reinforcement learning code is available on GitHub (https://github.com/efouragnan/RL_models). For any further enquiries regarding the data, please contact Elsa Fouragnan.

Acknowledgements

The authors thank Dr Maryann Noonan and Dr Marco Wittmann for fruitful discussions on the probabilistic task; research assistants Ema Darrieutort and Joshua Marquez for their work on the safety data, all study participants for taking part in the study, and the Brain Research & Imaging Centre (BRIC) MRI radiographers for their help with scanning. This research was supported by a UKRI Medical Research Council Future Leaders Fellowship grant and BBSRC grant (MR/T023007/1 and BB/Y001494/1) (to E.F.F.). Scanning for this study was supported by the Brain Research & Imaging Centre (BRIC). Effort for N.S.P. was supported by the National Institutes of Mental Health (U01 MH123427) and US Dept of Veterans Affairs (I50 RX002864). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the funders.

Author Contributions

S.N.Y and E.F.F. conceived this research and designed the study. N.S.P. advised on ultrasound safety of subcortical regions. J.R. advised on MRI acquisition. N.B. and M.R. advised on data quality and analysis. S.N.Y., E.B., M.L., N.B., and E.F.F. acquired data for the study. S.N.Y. and E.F.F. analysed the data. S.N.Y., N.B., M.R., and E.F.F. wrote the manuscript with input from all authors. All authors reviewed the final manuscript.

Competing Interests

The authors have no competing interests to declare.