



# High-frequency neuromodulation improves obsessive-compulsive behavior

Shrey Grover<sup>1</sup>, John A. Nguyen<sup>1</sup>, Vighnesh Viswanathan<sup>1</sup> and Robert M. G. Reinhart<sup>1,2,3,4</sup>✉

**Nearly one billion people worldwide suffer from obsessive-compulsive behaviors<sup>1,2</sup>, yet our mechanistic understanding of these behaviors is incomplete, and effective therapeutics are unavailable. An emerging perspective characterizes obsessive-compulsive behaviors as maladaptive habit learning<sup>3,4</sup>, which may be associated with abnormal beta-gamma neurophysiology of the orbitofrontal-striatal circuitry during reward processing<sup>5,6</sup>. We target the orbitofrontal cortex with alternating current, personalized to the intrinsic beta-gamma frequency of the reward network, and show rapid, reversible, frequency-specific modulation of reward- but not punishment-guided choice behavior and learning, driven by increased exploration in the setting of an actor-critic architecture. Next, we demonstrate that chronic application of the procedure over 5 days robustly attenuates obsessive-compulsive behavior in a non-clinical population for 3 months, with the largest benefits for individuals with more severe symptoms. Finally, we show that convergent mechanisms underlie modulation of reward learning and reduction of obsessive-compulsive symptoms. The results contribute to neurophysiological theories of reward, learning and obsessive-compulsive behavior, suggest a unifying functional role of rhythms in the beta-gamma range, and set the groundwork for the development of personalized circuit-based therapeutics for related disorders.**

Obsessive-compulsive behaviors are highly prevalent in the general population and cause significant distress to the individual<sup>1,2</sup>. Due to their presence in multiple psychiatric conditions, such as obsessive-compulsive disorder (OCD), these behaviors are increasingly identified as a core dimension for psychiatric classification<sup>7</sup>. Existing treatments for the heterogenous obsessive-compulsive symptoms produce disappointing outcomes and are accompanied by side effects that weaken compliance<sup>8,9</sup>. With non-invasive neuromodulation (for example, transcranial alternating current stimulation, tACS), interventions can be personalized according to individual neurophysiological dynamics<sup>10</sup>. Such personalized interventions have been proposed to improve treatment outcomes for obsessive-compulsive behaviors<sup>11</sup>.

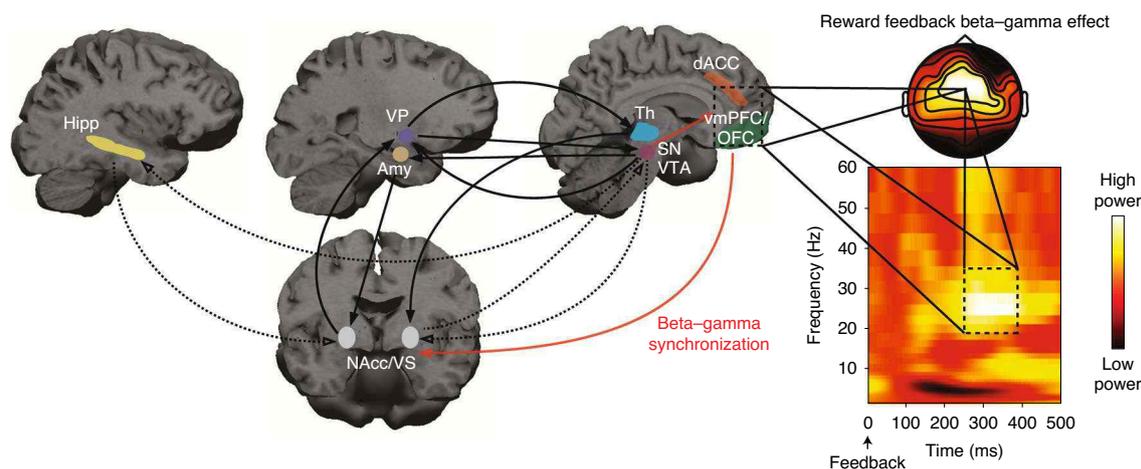
Obsessive-compulsive behaviors are considered outcomes of maladaptive habit learning due to abnormalities in a frontostriatal network engaged in repetitive execution of learned behaviors<sup>3,4</sup>. Previous studies have associated these behaviors with excessive habit learning as a function of reward during reinforcement learning<sup>12</sup>, along with abnormal connectivity between the medial orbitofrontal cortex (OFC) and the ventral striatum<sup>3,4,12–14</sup>. Interestingly, a parallel line of work has proposed functionally relevant oscillatory activity in the same network during reward learning. Novel, unexpected or

improbable reward, but not punishment, feedback elicits enhanced activity in the high beta–low gamma frequency range over frontal sites in human magneto- and electroencephalography (MEG/EEG, 20–35 Hz; Fig. 1)<sup>5,6,15,16</sup>. This effect is thought to be a motivational value signal originating from the OFC<sup>5,6</sup> that is transmitted to the ventral striatum to link preceding stimulus–response mapping to consequent rewards<sup>5,17</sup>. Modulation of these rhythms in the OFC during reward processing might influence the learning mechanisms facilitating transition to habitual behavior, thereby normalizing obsessive-compulsive behaviors.

Across two experiments, we used personalized, non-invasive, high-definition tACS (HD-tACS) to establish the functional contribution of OFC beta-gamma rhythms in reward learning and examined its potential in normalizing obsessive-compulsive behaviors. First (experiment 1), we sought to causally demonstrate that beta-gamma neuromodulation selectively influences reward- but not punishment-guided learning. We determined the intrinsic reward-related beta-gamma frequency of participants performing a reinforcement learning task and used HD-tACS to modulate these rhythms in the OFC during the task. Additionally, given that beta and gamma rhythms are proposed to map rewards with preceding events<sup>5,17</sup>, they may facilitate transition to habitual behavior, which can be reflected in exploitative decision-making<sup>18</sup>. However, reward-related beta-gamma activity has also been associated with low tonic prefrontal dopamine<sup>19</sup>, which is further associated with exploratory decision-making<sup>20</sup>. We used reinforcement learning models to study these candidate computational mechanisms during beta-gamma neuromodulation. Second (experiment 2), we examined the impact of chronic beta-gamma OFC modulation over 5 days on obsessive-compulsive characteristics in a non-clinical population. Similar to experiment 1, personalized neuromodulation frequencies were determined during a previous session while participants performed the reinforcement learning task. Obsessive-compulsive behaviors were examined over a period of 3 months. Finally, to identify common underlying mechanisms, we examined how neuromodulation-induced changes in reward learning and in obsessive-compulsive symptoms across both experiments are associated with intrinsic beta-gamma dynamics.

To examine the functional role of beta-gamma rhythms in reward learning, we used a between-participants, double-blind, double-sham-controlled design in experiment 1 involving a monetary reinforcement learning task<sup>21</sup> (Fig. 2a). Sixty participants were randomly assigned to either the passive sham, active control (alpha, 10 Hz) or personalized beta-gamma (mean, 26.6 Hz  $\pm$  3.3 s.d.) modulation group. Personalized beta-gamma frequencies were determined through a pre-experiment EEG recording while participants learned to associate stimuli with monetary rewards and losses. Electrical field

<sup>1</sup>Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA. <sup>2</sup>Center for Systems Neuroscience, Boston University, Boston, MA, USA. <sup>3</sup>Cognitive Neuroimaging Center, Boston University, Boston, MA, USA. <sup>4</sup>Center for Research in Sensory Communication and Emerging Neural Technology, Boston University, Boston, MA, USA. ✉e-mail: [rmgr@bu.edu](mailto:rmgr@bu.edu)



**Fig. 1 | Model integrating beta-gamma activity with reward learning circuitry.** Arrows show anatomical relationships between components of the reward and learning circuitry. Dashed black arrows show the loop of activity between the hippocampus (Hipp), nucleus accumbens (NAcc) and ventral tegmental area (VTA) put forward by Axmacher et al.<sup>38</sup>, with solid black arrows showing other major connections involved in learning and motivation<sup>39</sup>. The positive reward feedback beta-gamma effect shown to the right is theorized to derive from generators in the orbitofrontal cortex (OFC) or ventromedial prefrontal cortex (vmPFC)<sup>5,6</sup>, to be modulated by the substantia nigra (SN) and VTA (short red arrow), and to synchronize with beta-gamma rhythms in the ventral striatum (VS), providing a key mechanism by which the communication between frontostriatal regions can result in the transmission of a motivational value signal to the reward circuit in learning or decision-making contexts (long red arrow)<sup>5</sup>. Amy, amygdala; dACC, dorsal anterior cingulate cortex; Th, thalamus; VP, ventral pallidum.

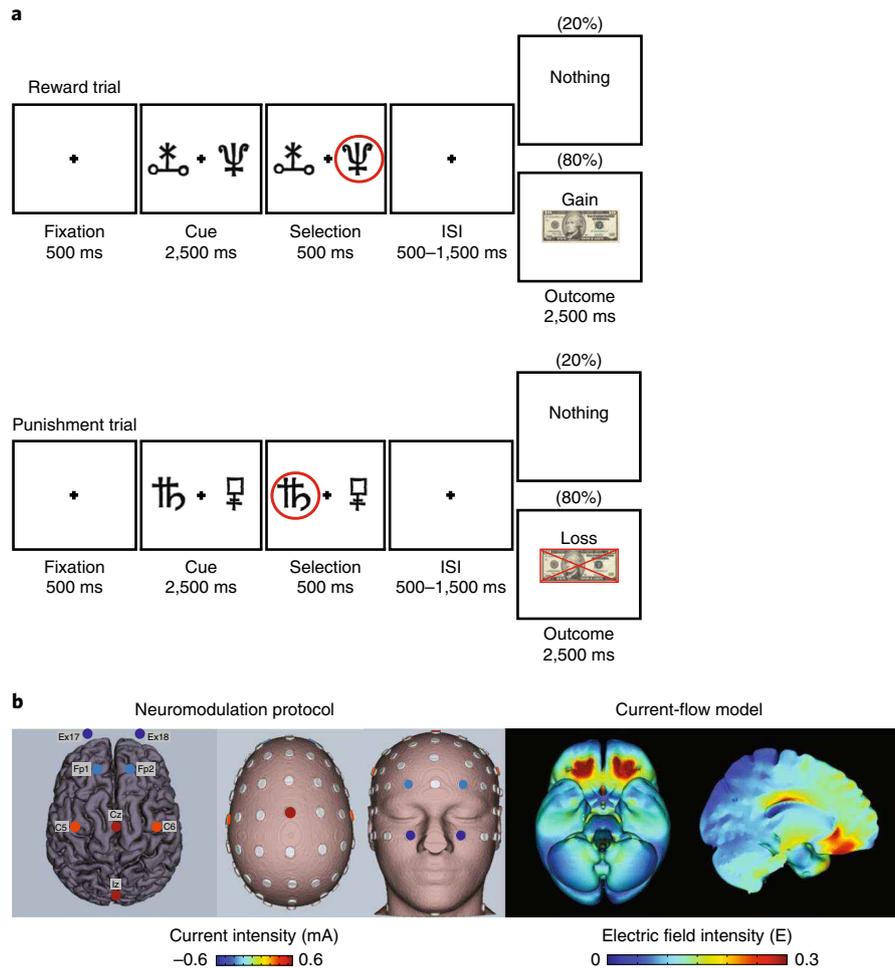
modeling was used to target the medial OFC (Fig. 2b). All participants performed a 30 min block of the task before, during and after modulation, and their choice behavior during each period was analyzed separately for reward and punishment trials.

Optimal choice behavior showed a significant interaction effect of the modulation group (sham, alpha, beta-gamma), valence (reward, punishment), and time (pre-modulation, modulation, post-modulation) (Fig. 3). Differences in choice behavior were observed when comparing beta-gamma with sham, and beta-gamma with alpha, but not when comparing alpha with sham, which suggests that beta-gamma modulation alone changed behavior. Comparing optimal choice behavior across groups during the modulation period, we found significantly impaired behavior during reward trials in the beta-gamma group. No differences across groups were observed for punishment trial behavior. In the beta-gamma group, choice behavior during reward trials was significantly impaired during the modulation period relative to the pre-modulation period but recovered in the post-modulation period, which suggests that the effects were acutely sensitive to concurrent beta-gamma HD-tACS (Fig. 3c). No differences were observed in punishment trial behavior over time. Control analyses in the alpha and sham groups did not show any differences over time in reward or punishment trials (all  $P > 0.468$ , all  $n = 20$ ). The three groups did not differ in the pre-modulation performance for reward (all  $P > 0.369$ , all  $n = 40$ ) or punishment trials (all  $P > 0.604$ , all  $n = 40$ ); this rules out potential between-participant confounding effects related to performance fluctuations at baseline. Together, these results demonstrate that frequency-personalized, beta-gamma OFC neuromodulation can exert a strong, valence-specific influence on choice behavior during reward learning in a rapid and reversible manner.

Next, we used reinforcement learning models to identify the computational mechanisms affected by beta-gamma neuromodulation during reward learning. We fitted these models to each participant's choice behavior during reward trials. We used four models: a standard Q-learning algorithm (StandardQ)<sup>22</sup>; a standard Q-learning algorithm with a free parameter allowing differential weighting of positive and negative outcomes (WeightedQ)<sup>23</sup>; an

actor-critic algorithm (Actor-Critic)<sup>22,23</sup>; and a hybrid model incorporating both Q-learning and actor-critic learning (Hybrid)<sup>23,24</sup>. The Actor-Critic model had the best fit or least amount of information loss, as indicated by the lowest Akaike information criterion pooled across participants and conditions (StandardQ, 34,435.68; WeightedQ, 32,703.25; Actor-Critic, 28,817.58; Hybrid, 31,126.38). Repeated measures analysis of variance (ANOVA) of the Actor-Critic model parameters for the within-participants factor of time (pre-modulation, modulation, post-modulation) and the between-participants factor of group (sham, alpha, beta-gamma) showed a significant interaction effect on mean temperature ( $\beta$ ) (Extended Data Fig. 1). Follow-up ANOVAs showed a significant time  $\times$  group interaction for beta-gamma compared with alpha, and for beta-gamma compared with sham, but not for alpha compared with sham. These interactions were driven by an increase in temperature during the modulation period, which rapidly recovered once neuromodulation was switched off, thus mirroring the effects of beta-gamma HD-tACS on reward choice behavior (Fig. 3). No other model parameters showed significant effects. To summarize, beta-gamma OFC neuromodulation acutely impaired reward-guided choice behavior by increasing an exploratory tendency, perhaps through the addition of decision noise or indeterminacy to the neural computations involved in choice.

After establishing that beta-gamma OFC HD-tACS modulates reward learning, we examined its effects on obsessive-compulsive behaviors. In experiment 2, we used a between-participants, double-blind, active-sham-controlled design. A non-clinical group of 64 participants with obsessive-compulsive symptoms in the subclinical and clinical range was studied. For each participant, we determined the endogenous reward-sensitive beta-gamma peak frequency using an abbreviated version of the learning task with concurrent EEG. Participants were assigned to either the personalized beta-gamma group (mean 27.1 Hz  $\pm$  3.6 s.d.) or the active control group undergoing alpha (10 Hz) modulation (that is, active control/alpha). HD-tACS, targeted at the medial OFC, was applied for 30 min each day for 5 consecutive days. We quantified obsessive-compulsive behaviors using the Obsessive-Compulsive Inventory-Revised (OCI-R)<sup>25</sup>, the most commonly used self-report



**Fig. 2 | Monetary reinforcement learning task and orbitofrontal neuromodulation protocol.** **a**, Participants chose one of two abstract visual stimuli and later observed the outcome. Following a jittered interstimulus interval (ISI), the outcome depending on the trial type was shown. In reward trials, one stimulus was associated with an 80% probability of winning US\$10 and a 20% probability of winning nothing, and the other stimulus had the opposite probability structure. In punishment trials, one stimulus was associated with an 80% probability of losing US\$10 and a 20% probability of losing nothing, and the other stimulus had the opposite probability structure. **b**, Personalized orbitofrontal neuromodulation protocol and current-flow models on three-dimensional reconstructions of the cortical surface. The location and current intensity value of each electrode are shown.

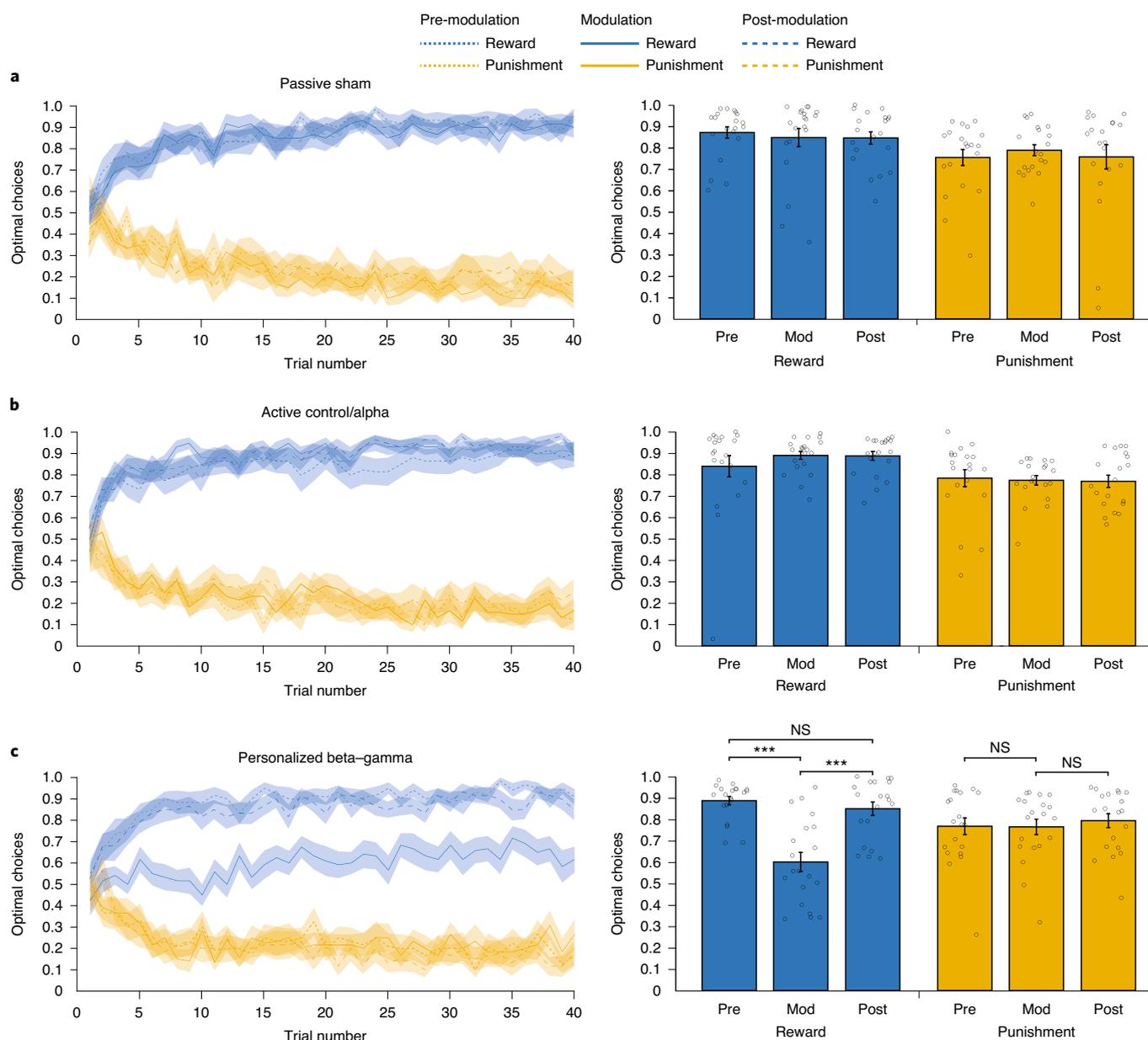
scale in OCD research<sup>26,27</sup>. The baseline measurement was taken prior to HD-tACS on the first day. Follow-up measurements were taken on the fifth day (the last day of HD-tACS), and at 1 month, 2 months and 3 months after the last day of HD-tACS.

Beta-gamma modulation induced a robust and long-lasting improvement in obsessive-compulsive behaviors. A significant group  $\times$  time interaction was observed in the OCI-R total score. In the beta-gamma group ( $n=32$ ), Bonferroni-corrected pairwise comparisons showed significant OCI-R reductions between pre-modulation baseline and all post-modulation periods (Fig. 4a, left panel): on the last intervention day ( $P=0.042$ ), and at the post-intervention timepoints of 1 month ( $P=0.013$ ), 2 months ( $P=0.002$ ) and 3 months ( $P=0.003$ ). On average, the OCI-R total score reduced by nearly 28% over 3 months. No significant changes in the OCI-R score were observed in the alpha group. Thus, beta-gamma neuromodulation induced frequency-specific improvements in obsessive-compulsive symptoms that could not be attributed to stimulation artifacts.

Beta-gamma modulation selectively improved several OCI-R subscales related to compulsivity (Fig. 4a, right panels). There were significant group  $\times$  time interactions for hoarding, ordering and obsessing behaviors. Within the beta-gamma group, main effects of

time were observed for hoarding, checking, ordering and washing behaviors. Follow-up Bonferroni-corrected pairwise comparisons in the beta-gamma group revealed rapid reductions in the OCI-R hoarding and ordering subscales by the last day of intervention. These reductions were sustained, to the post-intervention timepoint of 3 months. Of note, significant reductions were also observed in checking for up to 2 months. The alpha group showed no main effects of time and no pairwise differences at any timepoint relative to baseline. Pearson correlations between all time intervals in the alpha group showed good to excellent reliability for total and subscale scores (0.526 to 0.947, respectively; all  $P < 0.002$ , all  $n=32$ ), consistent with previous test-retest reliability studies of OCI-R in clinical and subclinical OCD populations<sup>25,28</sup>.

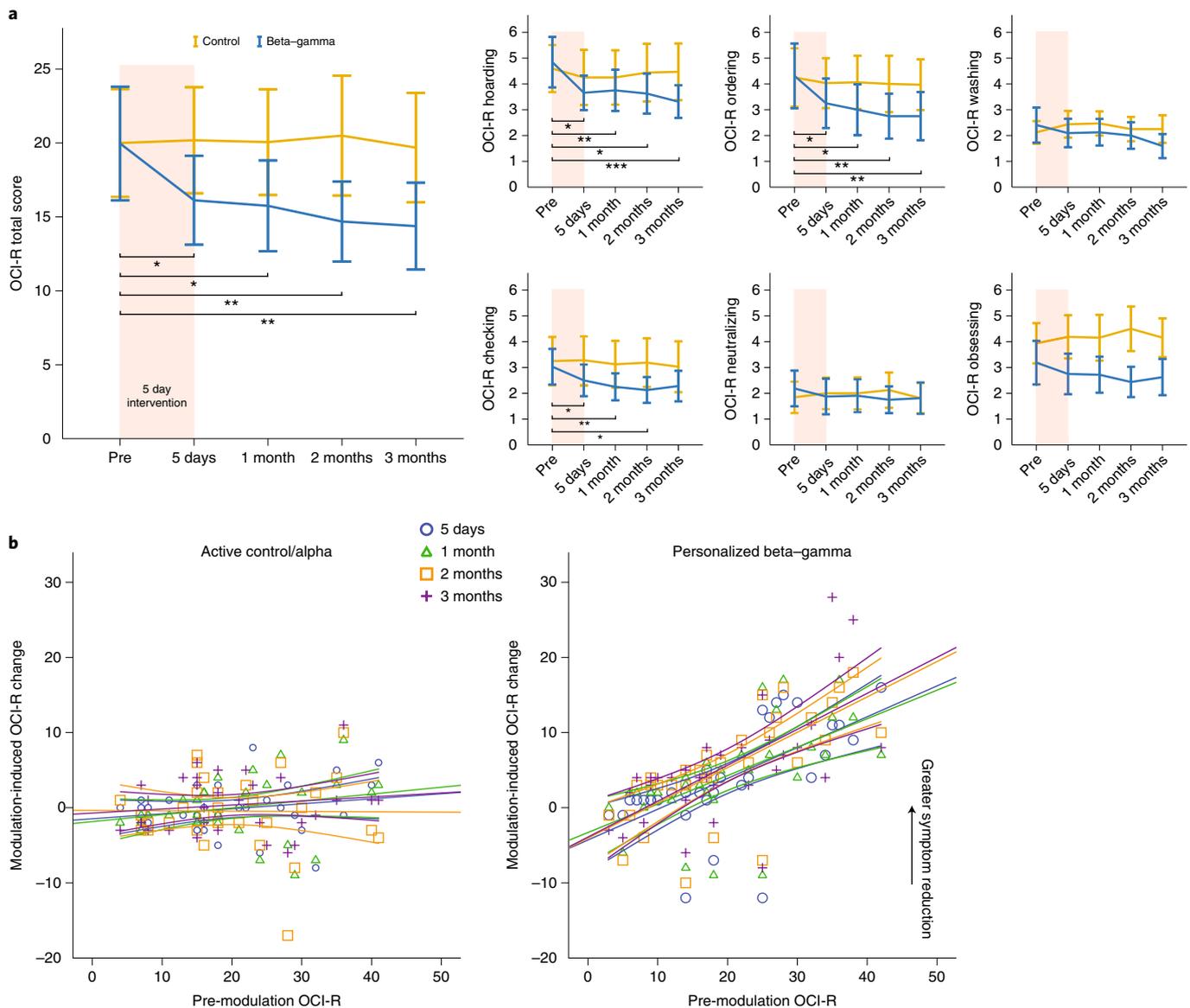
Participants with higher OCI-R scores at baseline showed greater reductions in obsessive-compulsive behaviors after beta-gamma modulation. We performed participant-wise correlations in the beta-gamma group between the total OCI-R score at baseline and the change in this score at each post-modulation timepoint. After correcting for multiple comparisons, significant correlations were found at each timepoint; this indicates that individuals with higher baseline ratings had the largest HD-tACS improvements (Fig. 4b). Conversely, individuals with relatively low OCI-R scores at baseline had no change



**Fig. 3 | Results of experiment 1, the monetary reinforcement learning task. a–c,** Observed optimal behavioral choices for reward (blue) and punishment (gold) during the pre-modulation (Pre, dotted), modulation (Mod, solid), and post-modulation (Post, dashed) blocks for the passive sham (a), active control/alpha (b) and personalized beta-gamma (c) groups. A significant modulation group  $\times$  valence  $\times$  time interaction was observed ( $F_{4,114} = 4.337$ ,  $P = 0.003$ ,  $\eta_p^2 = 0.132$ ,  $n = 60$ ). The interaction effect was present when comparing beta-gamma with sham ( $F_{2,76} = 3.643$ ,  $P = 0.036$ ,  $\eta_p^2 = 0.087$ ,  $n = 40$ ) and beta-gamma with alpha ( $F_{2,76} = 7.162$ ,  $P = 0.001$ ,  $\eta_p^2 = 0.159$ ,  $n = 40$ ), but not when comparing alpha with sham ( $F_{2,76} = 1.286$ ,  $P = 0.282$ ,  $n = 40$ ). During the modulation period, a significant group effect was observed for reward trials ( $F_{2,57} = 17.735$ ,  $P = 1 \times 10^{-6}$ ,  $\eta_p^2 = 0.384$ ,  $n = 60$ ), with beta-gamma modulation driving the effect (sham versus beta-gamma,  $F_{1,38} = 16.170$ ,  $P = 2.65 \times 10^{-4}$ ,  $\eta_p^2 = 0.299$ ,  $n = 40$ ; alpha versus beta-gamma,  $F_{1,38} = 35.997$ ,  $P = 5.69 \times 10^{-7}$ ,  $\eta_p^2 = 0.486$ ,  $n = 40$ ; sham versus alpha,  $F_{1,38} = 0.805$ ,  $P = 0.375$ ,  $n = 40$ ), but no differences were observed for punishment trials across groups (all  $F < 0.291$ , all  $P > 0.593$ , all  $n = 40$ ). In the beta-gamma group (c, right panel), there was a significant effect of time on reward trials ( $F_{2,38} = 20.159$ ,  $P = 2 \times 10^{-6}$ ,  $\eta_p^2 = 0.515$ ,  $n = 20$ ), with the modulation period showing significant differences compared with both pre-modulation ( $F_{1,19} = 29.533$ ,  $P = 3 \times 10^{-5}$ ,  $\eta_p^2 = 0.609$ ,  $n = 20$ ) and with post-modulation ( $F_{1,19} = 23.012$ ,  $P = 1.25 \times 10^{-4}$ ,  $\eta_p^2 = 0.548$ ,  $n = 20$ ) periods. No differences were observed between pre-modulation and post-modulation periods in reward trials ( $F_{1,19} = 0.822$ ,  $P = 0.376$ ,  $n = 20$ ). Beta-gamma modulation did not affect choice behavior in punishment trials over time (all  $F < 0.370$ , all  $P > 0.550$ ,  $ns = 20$ ). Mixed ANOVAs used the within-participants factors of valence (reward, punishment) and time (pre-modulation, modulation, post-modulation), and the between-participants factor of group (sham, alpha, beta-gamma). Error bars and shaded error bands,  $\pm 1$  s.e.m. \*\*\* $P < 0.001$ . NS, not significant.

or had slightly increased symptoms after repeated beta-gamma modulation. However, the number of low-scoring participants was a clear minority (6 of 32). The overwhelming majority of participants (26 of 32) had modulation-induced improvements, especially those who reported more severe obsessive-compulsive behaviors.

In addition to the total OCI-R score, subscale scores at baseline also predicted modulation-induced improvements at various timepoints. Significant correlations that survived  $P$  value correction for multiple comparisons ( $P < 0.0008$ ) and were present for at least three out of four post-modulation timepoints were



**Fig. 4 | Results of experiment 2, change in obsessive-compulsive symptoms after HD-tACS. a**, Mean OCI-R total score (left panel) and mean OCI-R subscale scores (right panels) with 95% confidence intervals (with mean at center of error bars) shown for each modulation group and timepoint. The shaded region shows the 5 day intervention period. For total OCI-R (left panel), there was a significant group  $\times$  time interaction ( $F_{4,248} = 6.748$ ,  $P = 2.43 \times 10^{-4}$ ,  $\eta_p^2 = 0.098$ ,  $n = 64$ ). There was a main effect of time on OCI-R in the beta-gamma group ( $F_{4,124} = 10.278$ ,  $P = 5.5 \times 10^{-5}$ ,  $\eta_p^2 = 0.249$ ,  $n = 32$ ), and no effect for the control/alpha group ( $F_{4,124} = 0.307$ ,  $P = 0.80$ ,  $n = 32$ ). See main text for statistics of pairwise comparisons. For subscales, significant group  $\times$  time effects were evident for hoarding ( $F_{4,248} = 2.886$ ,  $P = 0.030$ ,  $\eta_p^2 = 0.044$ ,  $n = 64$ ), ordering ( $F_{4,248} = 4.234$ ,  $P = 0.004$ ,  $\eta_p^2 = 0.064$ ,  $n = 64$ ) and obsessing ( $F_{4,248} = 2.817$ ,  $P = 0.032$ ,  $\eta_p^2 = 0.043$ ,  $n = 64$ ). The beta-gamma modulation group showed significant main effects of time for hoarding ( $F_{4,124} = 8.169$ ,  $P = 1.16 \times 10^{-4}$ ,  $\eta_p^2 = 0.209$ ,  $n = 32$ ), checking ( $F_{4,124} = 4.827$ ,  $P = 0.008$ ,  $\eta_p^2 = 0.135$ ,  $n = 32$ ), ordering ( $F_{4,124} = 9.8$ ,  $P = 3.9 \times 10^{-5}$ ,  $\eta_p^2 = 0.24$ ,  $n = 32$ ) and washing ( $F_{4,124} = 3.029$ ,  $P = 0.043$ ,  $\eta_p^2 = 0.089$ ,  $n = 32$ ). Mixed ANOVAs used the within-participants factor of time and the between-participants factor of group (control/alpha, beta-gamma). Follow-up Bonferroni-corrected pairwise comparisons (right panels) in the beta-gamma group ( $n = 32$ ) revealed a rapid reduction in hoarding ( $P = 0.02$ ) and ordering ( $P = 0.012$ ) by the last day of intervention, relative to baseline. These reduced scores remained reduced or continued to drop at 1 month (hoarding,  $P = 0.009$ ; ordering,  $P = 0.01$ ), 2 months (hoarding,  $P = 0.039$ ; ordering,  $P = 0.002$ ) and 3 months (hoarding,  $P = 1.9 \times 10^{-5}$ ; ordering,  $P = 0.003$ ). Of note, relative to baseline, there were significant reductions in checking on day 5 ( $P = 0.035$ ), at 1 month ( $P = 0.008$ ) and at 2 months ( $P = 0.015$ ) after intervention. The alpha group ( $n = 32$ ) showed no main effects of time (all  $F < 1.359$ , all  $P > 0.257$ ) and no pairwise differences at any timepoint relative to baseline (all  $P > 0.228$ ). **b**, Scatter plots with 95% confidence intervals of individual participant total OCI-R scores obtained before the intervention (pre-modulation) shown against the modulation-induced change in OCI-R (that is, baseline minus post-modulation measurement) for each post-intervention timepoint and for each modulation group. The beta-gamma group ( $n = 32$ ) showed significant Pearson correlations (two-tailed, corrected for multiple comparisons) on day 5 ( $r_{32} = 0.624$ ,  $P = 1.35 \times 10^{-4}$ ), 1 month ( $r_{32} = 0.602$ ,  $P = 2.71 \times 10^{-4}$ ), 2 months ( $r_{32} = 0.712$ ,  $P = 5 \times 10^{-6}$ ) and 3 months ( $r_{32} = 0.651$ ,  $P = 5.6 \times 10^{-5}$ ). \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

observed for hoarding (all  $r_{32} > 0.638$ , all  $P < 8.7 \times 10^{-5}$ , all  $n = 32$ ), checking (all  $r_{32} > 0.628$ , all  $P < 1.2 \times 10^{-4}$ , all  $n = 32$ ), ordering (all  $r_{32} > 0.590$ , all  $P < 3.8 \times 10^{-4}$ , all  $n = 32$ ) and washing (all  $r_{32} > 0.590$ ,

all  $P < 3.8 \times 10^{-4}$ , all  $n = 32$ ). By contrast, analyses of the same relationships for total OCI-R (all  $r_{32} < 0.237$ , all  $P > 0.192$ , all  $n = 32$ ) and subscales (all  $r_{32} < 0.554$ , all  $P > 0.001$ , all  $n = 32$ ) in the active

control/alpha group did not reach significance at the corrected alpha level and the large majority of correlations from the control group were not significant even at an alpha level of  $P < 0.05$  (23 out of 28, all  $r_{32} < 0.336$ , all  $P > 0.06$ , all  $n = 32$ ), thus confirming the improvements to be frequency specific in the beta–gamma range. The intervention was seen to be most effective for non-clinical individuals with more severe compulsivity characteristics.

Across the two experiments, we observed impairments in reward learning and reduction in obsessive–compulsive symptoms due to beta–gamma OFC modulation. Next, we devised a series of complementary analyses across both experiments to confirm that changes in reward learning and obsessive–compulsive behaviors were related to each other and similarly associated with OFC beta–gamma neurophysiology. If obsessive–compulsive behaviors arise from maladaptive habit learning, then individuals with more severe obsessive–compulsive characteristics should have higher optimal choice behavior during reward trials. To test this prediction, we divided participants in experiment 2 into two groups based on their baseline OCI-R score (low versus high), collapsed across neuromodulation groups, and analyzed their reward choice behavior during the abbreviated, pre-experiment reinforcement learning task. The group with high OCI-R scores had significantly higher behavioral accuracy (mean  $\pm$  s.e.m.,  $0.91 \pm 0.02$ ,  $n = 34$ ) than the low OCI-R group ( $0.83 \pm 0.02$ ,  $n = 30$ ;  $F_{1,62} = 6.759$ ,  $P = 0.012$ ,  $\eta_p^2 = 0.098$ ,  $n = 64$ ). Participant-wise regression analysis revealed that individuals with high reward choice behavior were also those with high OCI-R scores ( $r_{64} = 0.494$ ,  $P = 3.4 \times 10^{-5}$ ,  $n = 64$ ). Thus, high symptom severity was associated with better learning from rewards, consistent with previous work linking compulsivity and excessive reward-based habit learning<sup>12</sup>.

Next, we examined the relationship between reward choice behavior and the change in the intrinsic beta–gamma power in response to rewarding feedback relative to punishing feedback. We computed the beta–gamma effect during the pre-experiment EEG in experiment 1 and correlated it with the choice behavior during the pre-modulation periods. A significant participant-wise negative correlation ( $r_{60} = -0.467$ ,  $P = 1.7 \times 10^{-4}$ ,  $n = 60$ ) was observed, which suggests that individuals who exhibited a smaller beta–gamma effect in response to rewarding feedback consistently made better choices during reward trials.

Optimal reward choice behavior was associated with a smaller beta–gamma effect (in experiment 1) and higher obsessive–compulsive symptom severity (in experiment 2). This implies that individuals with higher OCI-R scores should show a smaller beta–gamma effect in their pre-experiment EEG. We compared the beta–gamma effect between individuals with high and low OCI-R total scores at baseline. An ANOVA using the within-participants factor of valence (reward versus punishment) and the between-participants factor of OCI-R score (high versus low) indicated a significant valence  $\times$  OCI-R interaction ( $F_{1,62} = 9.462$ ,  $P = 0.003$ ,  $\eta_p^2 = 0.132$ ,  $n = 64$ ). By parsing the interaction, we observed that the high OCI-R group exhibited no beta–gamma effect, in agreement with our prediction, given that power values relative to baseline on reward trials (mean  $\pm$  s.e.m.,  $0.013 \pm 0.010$  dB) and on punishment trials ( $0.0003 \pm 0.008$  dB) did not significantly differ ( $F_{1,33} = 0.954$ ,  $P = 0.336$ ,  $\eta_p^2 = 0.028$ ,  $n = 34$ ). By contrast, individuals who scored relatively low on the OCI-R scale showed a significant beta–gamma effect ( $F_{1,29} = 20.712$ ,  $P = 8.8 \times 10^{-5}$ ,  $\eta_p^2 = 0.417$ ,  $n = 30$ ) due to a large increase in beta–gamma power on reward trials ( $0.074 \pm 0.011$  dB) relative to punishment trials ( $0.008 \pm 0.013$  dB). Critically, these observations suggest that both obsessive–compulsive behaviors and reward-guided choice behavior during learning are similarly associated with rhythms in the beta–gamma range.

As previously discussed, modulation gains were stronger for individuals with more severe obsessive–compulsive characteristics (Fig. 4b). Given that obsessive–compulsive characteristics were

associated with better choice performance in reward trials, individuals with more optimal reward choice behavior at baseline should experience stronger impairments during beta–gamma neuromodulation. Indeed, a significant negative correlation was observed between reward choice behavior during the pre-modulation and modulation periods across participants in the beta–gamma group in experiment 1 ( $r_{20} = -0.578$ ,  $P = 0.008$ ,  $n = 20$ ). Together, these results establish parallel effects of neuromodulation in both experiments and suggest that convergent mechanisms underlie the regulation of reward-guided choice behavior and obsessive–compulsive behaviors through beta–gamma neuromodulation.

Activity in the beta–gamma range is considered a key mechanism for coupling frontostriatal areas during reward learning<sup>5</sup>. These rhythms are thought to transmit motivational value signals within the reward network to assist learning (Fig. 1)<sup>5,17</sup>. In light of these theories, it might seem surprising that modulation of these rhythms in the OFC in the current study caused impairments, and not enhancements, in optimal choice behavior. A facilitatory role of beta–gamma rhythms in learning may still emerge if both OFC and ventral striatum are simultaneously modulated. For synchronized circuits, neuromodulation of a single node in the circuit may not be sufficient to provide additional benefits<sup>10</sup>. Development of temporal interference-based subcortical neuromodulation<sup>29</sup> for human studies, paired with OFC modulation using HD-tACS may be key to empirically test this possibility. Another possibility is that neuromodulation may have saturated the intrinsic beta–gamma activity in the OFC, which would leave little room for reward-induced modulations for behavioral adaptation. Although more work is needed to test these hypotheses, our observations of impaired choice behavior align with another perspective that takes into account the role of dopamine circuitry in the prefrontal cortex. Specifically, reward-related beta–gamma activity is stronger in individuals with reduced tonic prefrontal dopamine<sup>19</sup>. Furthermore, reduction in OFC dopamine levels is associated with increased exploratory behavior<sup>20</sup>. Examination of choice behavior through reinforcement learning models supports this perspective. Using the actor-critic framework, an increase was detected in exploration during decision-making while participants underwent beta–gamma neuromodulation. It is possible that beta–gamma neuromodulation increased local beta–gamma activity and reduced dopamine levels in the OFC to increase exploratory behavior. These modulations may be accompanied by changes in the frontostriatal connectivity and striatal dopamine levels<sup>20</sup>. The increase in exploration could also be due to added noise in maintaining or transmitting the identity information of the rewarding stimuli<sup>30</sup>, or due to downstream effects in a broader network that controls exploratory behavior such as the locus coeruleus<sup>31,32</sup> or the motor cortex<sup>33</sup>. Future neuromodulation studies carried out concurrently with electrophysiology and neuroimaging will be needed to examine these neural processes that are potentially influenced by beta–gamma modulation during reward-guided choice behavior.

Compulsive behaviors are characterized as outcomes of dysfunctional habit learning<sup>4,12</sup>. In support of these theories, we observed greater severity of compulsive behaviors associated with more exploitative reward choice behavior. Chronic administration of HD-tACS over the medial OFC personalized to the intrinsic beta–gamma frequencies of the reward network decreased compulsive behaviors for at least 3 months. Together with prior evidence of an association of obsessive–compulsive behavior with habit learning<sup>4,12</sup>, these observations suggest that OFC neuromodulation may have improved obsessive–compulsive symptoms by promoting exploratory behavior. Indeed, reduced exploration during reward choice behavior has been observed in some pathologies of compulsivity such as alcohol use disorder<sup>34</sup>, and in animal models of compulsive behavior<sup>35</sup>. It is plausible that chronic OFC neuromodulation recovered the dysfunctional frontostriatal connectivity associated

with compulsivity characteristics<sup>4</sup>, and may have affected the striatum, which also exhibits beta frequency activity during habitual behavior<sup>36</sup>. These predictions can be better examined through concurrent neuromodulation and neuroimaging. At present, our results provide causal evidence for the involvement of the OFC in obsessive-compulsive behaviors in humans, and the frequency-specific effects of neuromodulation provide a mechanistic insight into the neurophysiology that can be targeted for symptom improvement. The personalized neuromodulation design leverages the neuroplastic properties of the brain to produce long-lasting effects via chronic HD-tACS administration<sup>37</sup>. Although we examined a subclinical population in the present study, the benefits observed provide motivation for further investigation of personalized neuromodulation as an experimental medicine for clinical obsessive-compulsive spectrum disorders and other conditions of frontostriatal dysfunction.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-01173-w>.

Received: 30 April 2020; Accepted: 10 November 2020;

Published online: 18 January 2021

### References

- Fullana, M. A. et al. Obsessions and compulsions in the community: prevalence, interference, help-seeking, developmental stability, and co-occurring psychiatric conditions. *Am. J. Psychiatry* **166**, 329–336 (2009).
- Ruscio, A. M., Stein, D. J., Chiu, W. T. & Kessler, R. C. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Mol. Psychiatry* **15**, 53–63 (2010).
- Burguière, E., Monteiro, P., Mallet, L., Feng, G. & Graybiel, A. M. Striatal circuits, habits, and implications for obsessive-compulsive disorder. *Curr. Opin. Neurobiol.* **30**, 59–65 (2015).
- Robbins, T. W., Vaghi, M. M. & Banca, P. Obsessive-compulsive disorder: puzzles and prospects. *Neuron* **102**, 27–47 (2019).
- Marco-Pallarés, J., Münte, T. F. & Rodríguez-Fornells, A. The role of high-frequency oscillatory activity in reward processing and learning. *Neurosci. Biobehav. Rev.* **49**, 1–7 (2015).
- Cohen, M. X., Wilmes, K. & van de Vijver, I. Cortical electrophysiological network dynamics of feedback learning. *Trends Cogn. Sci.* **15**, 558–566 (2011).
- Gillan, C. M., Fineberg, N. A. & Robbins, T. W. A trans-diagnostic perspective on obsessive-compulsive disorder. *Psychol. Med.* **47**, 1528–1548 (2017).
- Bloch, M. H., McGuire, J., Landeros-Weisenberger, A., Leckman, J. F. & Pittenger, C. Meta-analysis of the dose-response relationship of SSRI in obsessive-compulsive disorder. *Mol. Psychiatry* **15**, 850–855 (2010).
- Hollander, E., Kaplan, A., Allen, A. & Cartwright, C. Pharmacotherapy for obsessive-compulsive disorder. *Psychiatr. Clin. North Am.* **23**, 643–656 (2000).
- Reinhart, R. M. G. & Nguyen, J. A. Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nat. Neurosci.* **22**, 820–827 (2019).
- Cocchi, L. et al. Transcranial magnetic stimulation in obsessive-compulsive disorder: a focus on network mechanisms and state dependence. *Neuroimage Clin.* **19**, 661–674 (2018).
- Voon, V. et al. Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Transl. Psychiatry* **5**, e670 (2015).
- Remijnse, P. L. et al. Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **63**, 1225–1236 (2006).
- Kubota, Y. et al. Corticostriatal-limbic correlates of sub-clinical obsessive-compulsive traits. *Psychiatry Res. Neuroimaging* **285**, 40–46 (2019).
- Marco-Pallarés, J. et al. Human oscillatory activity associated to reward processing in a gambling task. *Neuropsychologia* **46**, 241–248 (2008).
- Cohen, M. X., Elger, C. E. & Ranganath, C. Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage* **35**, 968–978 (2007).
- Mas-Herrero, E., Ripollés, P., HajiHosseini, A., Rodríguez-Fornells, A. & Marco-Pallarés, J. Beta oscillations and reward processing: coupling oscillatory activity and hemodynamic responses. *Neuroimage* **119**, 13–19 (2015).
- Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **31**, 359–387 (2008).
- Marco-Pallarés, J. et al. Genetic variability in the dopamine system (dopamine receptor D4, catechol-O-methyltransferase) modulates neurophysiological responses to gains and losses. *Biol. Psychiatry* **66**, 154–161 (2009).
- Clarke, H. F. et al. Orbitofrontal dopamine depletion upregulates caudate dopamine and alters behavior via changes in reinforcement sensitivity. *J. Neurosci.* **34**, 7663–7676 (2014).
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045 (2006).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
- Gold, J. M. et al. Negative symptoms in schizophrenia result from a failure to represent the expected value of rewards: behavioral and computational modeling evidence. *Arch. Gen. Psychiatry* **69**, 129–138 (2012).
- Hernaus, D., Gold, J. M., Waltz, J. A. & Frank, M. J. Impaired expected value computations coupled with overreliance on stimulus-response learning in schizophrenia. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 916–926 (2018).
- Foa, E. B. et al. The obsessive-compulsive inventory: development and validation of a short version. *Psychol. Assess.* **14**, 485–496 (2002).
- Andersson, E. et al. Internet-based cognitive behaviour therapy for obsessive-compulsive disorder: a randomized controlled trial. *Psychol. Med.* **42**, 2193–2203 (2012).
- Simpson, H. B. et al. A randomized, controlled trial of cognitive-behavioral therapy for augmenting pharmacotherapy in obsessive-compulsive disorder. *Am. J. Psychiatry* **165**, 621–630 (2008).
- Hajcak, G., Huppert, J. D., Simons, R. F. & Foa, E. B. Psychometric properties of the OCI-R in a college sample. *Behav. Res. Ther.* **42**, 115–123 (2004).
- Grossman, N. et al. Noninvasive deep brain stimulation via temporally interfering electric fields. *Cell* **169**, 1029–1041.e16 (2017).
- McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y. & Schoenbaum, G. Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J. Neurosci.* **31**, 2700–2705 (2011).
- McClure, S. M., Gilzenrat, M. S. & Cohen, J. D. In *Advances in Neural Information Processing Systems 18* (eds Weiss, Y., Schölkopf, B. & Platt, J. C.) 867–874 (MIT Press, 2006).
- Cohen, J. D., McClure, S. M. & Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 933–942 (2007).
- Schreiner, D. C. & Gremel, C. M. Orbital frontal cortex projections to secondary motor cortex mediate exploitation of learned rules. *Sci. Rep.* **8**, 10979 (2018).
- Morris, L. S. et al. Biases in the explore-exploit tradeoff in addictions: the role of avoidance of uncertainty. *Neuropsychopharmacology* **41**, 940–948 (2016).
- Thompson, S. L. et al. Btd3 expression regulates compulsive-like and exploratory behaviors in mice. *Transl. Psychiatry* **9**, 222 (2019).
- Howe, M. W., Atallah, H. E., McCool, A., Gibson, D. J. & Graybiel, A. M. Habit learning is associated with major shifts in frequencies of oscillatory activity and synchronized spike firing in striatum. *Proc. Natl Acad. Sci. U. S. A.* **108**, 16801–16806 (2011).
- Vossen, A., Gross, J. & Thut, G. Alpha power increase after transcranial alternating current stimulation at alpha frequency ( $\alpha$ -tACS) reflects plastic changes rather than entrainment. *Brain Stimul.* **8**, 499–508 (2015).
- Axmacher, N. et al. Intracranial EEG correlates of expectancy and memory formation in the human hippocampus and nucleus accumbens. *Neuron* **65**, 541–549 (2010).
- Camara, E., Rodríguez-Fornells, A., Ye, Z. & Münte, T. F. Reward networks in the brain as captured by connectivity measures. *Front. Neurosci.* **3**, 350–362 (2009).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## Methods

**Participants.** A total of 128 participants (60 in experiment 1, 68 in experiment 2) signed informed consent forms, approved by the Boston University Institutional Review Board, and were paid. For experiment 1, data on all 60 participants recruited were analyzed (sham,  $n = 20$ , mean age, 26.6 years  $\pm$  5.8 s.d., 10 female, all right-handed; alpha,  $n = 20$ , mean age, 25.8 years  $\pm$  5.6 s.d., 10 female, all right-handed; beta-gamma,  $n = 20$ , mean age, 25.1 years  $\pm$  6.2 s.d., 11 female, all right-handed; two-tailed independent samples  $t$ -tests showed no between-group age differences, all  $t_{38} < 0.790$ , all  $P > 0.435$ , all  $n = 40$ ). For experiment 2, we sought a sample that reflected a wide range of scores on the OCI-R for correlational analysis and to determine the impact of the beta-gamma intervention on people with varying levels of obsessive-compulsive behavior severity. This approach was motivated by conceptualizations in dimensional psychiatry that emerged as Research Domain Criteria (RDoC)<sup>40</sup> and align with transdiagnostic perspectives of obsessive-compulsive spectrum disorders<sup>7</sup>. To this end, we sought a relatively equal distribution of low and high OCI-R scores across the spectrum, similar to previous research methods<sup>41</sup>. Thirty-four participants with relatively low OCI-R scores (0–17) and another 34 participants with relatively high OCI-R scores ( $\geq 18$ ) were recruited. Four participants were lost to attrition due to multiple testing sessions. The final group of 64 participants consisted of 30 people with low scores (active control/alpha,  $n = 15$ , mean age 23.9 years  $\pm$  4.0 s.d., 8 female, all right-handed; beta-gamma,  $n = 15$ , 25.3 years  $\pm$  3.8 s.d., 8 female, all right-handed; two-tailed independent samples  $t$ -test for age,  $t_{28} = 1.026$ ,  $P = 0.314$ ,  $n = 30$ ), and 34 people with high scores (active control/alpha,  $n = 17$ , 23.4 years  $\pm$  4.0 s.d., 9 female, all right-handed; beta-gamma,  $n = 17$ , 23.0 years  $\pm$  3.7 s.d., 9 female, all right-handed; two-tailed independent samples  $t$ -test on age,  $t_{32} = 0.327$ ,  $P = 0.745$ ,  $n = 34$ ). Independent samples two-tailed  $t$ -tests showed no significant age differences between active control and beta-gamma modulation groups ( $t_{62} = 0.491$ ,  $P = 0.625$ ,  $n = 64$ ). Of the 34 participants with high scores, 28 (active control/alpha,  $n = 14$ ; beta-gamma,  $n = 14$ ) scored higher than 20, which indicates the likely presence of OCD<sup>25</sup>.

All participants reported no metal implants in head, no implanted electronic devices, no history of neurological problems or head injury, no skin sensitivity, no claustrophobia, not being pregnant, no history of an Axis I psychiatric disorder, no current use of psychoactive medication, no substance abuse or dependence in the past 6 months, normal color vision, normal or corrected-to-normal visual acuity, and not being enrolled in any other research or treatment study for the duration of their study participation.

**Stimuli and procedures.** *Overview.* Experiment 1 was a randomized, between-participants, double-blind study using passive and active control procedures and pre-post within-participants measurements. Sixty participants were randomly assigned to either the passive sham, the active control (alpha, 10 Hz) or the personalized beta-gamma (mean, 26.6 Hz  $\pm$  3.3 s.d.) experimental group (20 participants each), with the use of a covariate adaptive randomization method to produce equal-sized groups balanced by sex and age<sup>42</sup>. All procedures for each experimental group were identical, with the exception of the HD-tACS protocol (passive sham, active control/alpha or personalized beta-gamma; Fig. 2). In addition to a passive sham control (the gold standard in the field of non-invasive neuromodulation) we instituted an active control procedure using alpha-rate HD-tACS to minimize potential confounding effects related to peripheral co-stimulation and to test for frequency specificity of HD-tACS effects<sup>10,43</sup>. For each experimental group, the duration of task performance was 90 min, and this consisted of three consecutive periods (pre-modulation, modulation, post-modulation; 30 min each) during which participants performed a classic monetary reinforcement learning task (Fig. 2a). The additional pre-post, within-participants measurements increased the rigor and reproducibility of the study. The pre-modulation period offered an additional baseline period from which to compare the effects of each intervention within groups, and provided the opportunity to evaluate the stability of the baseline data collected across groups. The post-modulation period enabled pre-post comparisons, evaluation of practice effects, and the duration of modulation effects to be examined. Behavioral data were analyzed before, during and after the 30 min application of HD-tACS in each experimental group (sham, alpha, beta-gamma).

Experiment 2 was a randomized, between-participants, double-blind study using an active control procedure (alpha-rate HD-tACS) and pre-post, within-participants measurements. Sixty-four participants were randomly assigned to either the active control (alpha, 10 Hz) or personalized beta-gamma (mean, 27.1 Hz  $\pm$  3.6 s.d.) experimental group (32 participants each) using a covariate adaptive randomization method to produce equal-sized groups balanced by sex, age and OCI-R score<sup>42</sup>. The procedures across experimental groups were identical, with the exception of modulation frequency (10 Hz versus personalized  $\sim$ 27 Hz; Fig. 2). In addition to the use of electrical field modeling and HD-tACS to help to focalize the stimulated area and minimize unwanted peripheral co-stimulation<sup>44–46</sup>, we also instituted an active control procedure that matched the personalized beta-gamma intervention in terms of estimated effects on extra-cranial tissue, shunting and peripheral co-stimulation<sup>10,43</sup>. All participants completed a baseline pre-modulation OCI-R assessment, followed by 5 consecutive days of alpha or beta-gamma HD-tACS (30 min each day). On the fifth day, after neuromodulation,

participants completed a post-modulation OCI-R assessment, followed by three longer-term post-modulation OCI-R assessments at 1 month, 2 months and 3 months after the last day (that is, day 5) of the intervention. OCI-R scores were analyzed for all five data collection timepoints (baseline, day 5, 1 month, 2 months and 3 months).

**Experimental task.** In experiment 1, participants performed three consecutive blocks (pre-modulation, modulation, post-modulation) of a well-validated monetary reinforcement learning task (Fig. 2a), coded in Psychtoolbox<sup>47</sup>, based on methods from ref. 21. Each block was approximately 30 min and consisted of three runs, with new pairs of stimuli in each run. During a run, participants performed 80 trials (40 reward, 40 punishment). Each trial began with fixation (500 ms), followed by the presentation of two novel visual stimuli (letters from the Agathodaimon font; 2,500 ms) to the left and right of fixation (relative stimulus position counterbalanced across trials). Participants were instructed to choose between stimuli to maximize payoffs. Upon making a selection using a button press from a handheld gamepad, a red circle appeared around the chosen stimulus (500 ms). Following a jittered interstimulus interval (500–1,500 ms), feedback was presented (2,500 ms). Each pair of stimuli was related to a pair of outcomes (in the reward trials, the two outcomes were +US\$10 or US\$0, and in the punishment trials they were –US\$10 or US\$0), with the two stimuli in each trial each having reciprocal probabilities of the two outcomes (80%/20% and 20%/80%, respectively). Reward and punishment trials were randomly interleaved within a run. To maximize payoffs, participants needed to learn stimulus–outcome contingencies via trial and error. Participants were told that their wins from a randomly selected run of trials would be their compensation for participation, but in actuality all participants left with the same fixed amount.

**Clinical instrument.** In experiment 2, participants' obsessive-compulsive characteristics were evaluated using the OCI-R<sup>25</sup>. The OCI-R is one of the most commonly used self-report assessments in obsessive-compulsive behavior research, worldwide<sup>26,27</sup>. It has been used to quantify obsessive-compulsive symptoms in various disorders, including OCD<sup>25</sup>. It is an 18-item unipolar self-report questionnaire of DSM-IV obsessive-compulsive symptoms that includes six subscales: washing, checking, neutralizing, obsessing, ordering and hoarding, and these subscales have been validated in clinical<sup>48</sup>, non-clinical<sup>49</sup>, and combined samples<sup>25</sup>. Each response is scored on a five-point scale from 'not at all' (coded as 0) to 'extremely' (coded as 4). The total score is the sum of the scores on all 18 items. People with OCD typically have a score of 21 or higher<sup>25</sup>. The OCI-R has good to excellent convergent and discriminant validity, test-retest reliability, and internal consistency in clinical and non-clinical samples<sup>25,48,49</sup>.

**HD-tACS.** The alternating current stimulation was delivered non-invasively using a nine-channel high-definition transcranial electrical-current stimulator (Soterix Medical). The HD-tACS procedure used eight sintered 12-mm-diameter Ag-AgCl electrodes, attached to custom high-definition plastic holders, filled with conductive gel, and embedded in a BrainCap (Brain Vision). Electrical field modeling using HD-Explore and HD-Targets (Soterix Medical) guided decision-making about where to place electrodes, with the goal of delivering focalized current to medial regions of the bilateral OFC. Figure 2b shows the modulation parameters, including the number of electrodes, electrode location and current intensity values (Fp1, –0.3 mA; Fp2, –0.3 mA; Cz, 0.6 mA; C5, 0.3 mA; C6, 0.3 mA; Ex17, –0.6 mA; Ex18, –0.6 mA; Iz, 0.6 mA). A bipolar sinusoidal alternating current was applied at either a participant-defined beta-gamma frequency (that is, personalized beta-gamma) or at 10 Hz (that is, active control/alpha). Modulation intensity (peak to peak) was set to 1.8 mA. Modulation was done for 30 min while participants performed the reinforcement learning task (experiment 1) or were seated at rest with eyes open, not engaged in any task (experiment 2). All participants confirmed that the modulation procedure was acceptable and did not induce painful skin sensations or phosphenes.

**Control procedures.** We implemented numerous procedures, in accordance with previously established methods<sup>10,50–52</sup>, to ensure that information about the experiments would not bias results. First, both experiments were between-participant in design to minimize learning and the transfer of effects across conditions. Second, for both experiments, within-participant, pre- and post-modulation assessment periods were implemented for each modulation group to reduce random noise such as confounding effects related to participant differences. In addition to comparisons afforded by the between-participant design, the addition of pre-post assessment periods allowed us to: (1) examine the effects of each type of modulation through time in the same group of participants; (2) evaluate the stability of baseline task performance across the different groups of participants; and (3) determine how long the modulation effects lasted.

Third, the experiments were sham-controlled. Experiment 1 used the standard passive sham control. The passive sham procedure followed the same procedure as the active beta-gamma modulation, but lasted only 30 s, and ramped up and down at the beginning and end of the 30 min period, to simulate the tingling sensation that participants commonly experience and then quickly adjust to during active modulation<sup>51</sup>. Additionally, experiments 1 and 2 utilized an active control<sup>10</sup>.

The parameters of the active control procedure were identical to that of beta-gamma modulation, except that alternating current was applied at a neighboring frequency band (alpha, 10 Hz). This was an especially rigorous control procedure because unlike the passive sham, HD-tACS during active control was applied to the same cortical target (OFC), at the full 1.8 mA intensity, for the full 30 min duration. The active control effectively minimizes unwanted effects associated with changes in extra-cranial tissue, shunting and peripheral co-stimulation, such as transretinal or transcutaneous stimulation<sup>45</sup>. The improved neuromodulation technology (HD-tACS) and electrical field modeling that we used have also been shown to minimize the effects of peripheral co-stimulation because these methods enhance anatomical targeting precision and produce more focal current distributions<sup>44,53</sup>.

Fourth, we used a double-blind procedure, in which a second experimenter set the mode (for example, active or sham) on the stimulator, but otherwise did not interact with the participant or with the experimenter who performed data collection. Fifth, we conducted recordings in a sound-attenuated, electrically shielded chamber to avoid line noise interference and reduce the possibility of participant-experimenter interaction during testing, which could have at least implicitly biased the participants.

Last, we confirmed that participants remained blind to the type of modulation that they received. After the experiment had ended (experiment 1) or each modulation test day had ended (experiment 2), we administered a safety questionnaire<sup>54</sup> and visual analog scale<sup>55</sup>, which included questions regarding attention, concentration, mood, vision, headache, fatigue and skin sensations under electrodes. Scores on these ratings were not significantly different across modulation groups (experiment 1, all  $F_{1,38} < 0.307$ , all  $P > 0.583$ , all  $n = 40$ ; experiment 2, all  $F_{1,62} < 1.229$ , all  $P > 0.272$ , all  $n = 64$ ). In addition, all participants were asked at the end of the experiments whether they could guess which modulation group they had been randomly assigned to. Participants' guesses were near chance (experiment 1, 31.67% hit rate; experiment 2, 53.13% hit rate).

**Data analysis. Pre-experiment EEG frequency tuning.** Dynamic systems theory predicts that systems with intrinsic periodic dynamics have preferred modulation frequencies and that neuronal modulation should be most effective when the external modulation frequency of weak periodic perturbations is closest to the brain network's preferred frequency<sup>56</sup>. To design a personalized HD-tACS intervention, we determined the peak frequency of each participant's endogenous positive feedback beta-gamma effect. Each participant completed a brief pre-study EEG recording while they performed 30 min of the reinforcement learning task. EEG was recorded using our standard procedures<sup>10</sup>. The EEG signal was collected from 64 electrodes arrayed according to the international 10–20 system using an ActiChamp active channel amplifier sampling at 1,000 Hz (Brain Vision). The right mastoid electrode served as the online reference. Horizontal eye position was monitored by recording the electro-oculogram from bipolar electrodes placed at the outer canthus of each eye, and vertical eye position and blinks were monitored with bipolar electrodes placed above and below the left orbit.

Offline, data were re-referenced to the average of the left and right mastoids. Broad segments were cut from  $-1,000$  ms before to  $2,000$  ms after feedback onset to eliminate edge artifacts induced by wavelet convolution from the critical middle of this window. Visual inspection was used to remove large muscle artifacts, followed by independent component analysis to remove blink and noise artifacts<sup>57</sup>. The data were converted to current source density (CSD)<sup>58</sup> to improve topographical localization<sup>59</sup>. Total power was calculated by convolving the CSD-EEG time series in each segment with a set of complex Morlet wavelets, defined with a Gaussian envelope using a constant ratio ( $\sigma_f = f/7$ ) and a wavelet duration ( $6\sigma_t$ ), in which  $f$  is the center frequency and  $\sigma_t = 1/(2\pi\sigma_f)$ . Frequencies ranged from 1 to 60 in 0.5 Hz steps. Following single-trial EEG spectral decomposition, the magnitude (that is, length) of the complex number vectors was extracted, squared and averaged, yielding a measure of intertrial total power for a given frequency, timepoint and electrode. Power was decibel normalized, in which for each electrode and frequency, the average power during an interval of  $-100$  to  $0$  ms before feedback onset served as the baseline activity. Analysis focused on the beta-gamma band frequencies (20–35 Hz) at electrode Fz during a conservatively broad temporal window (200–500 ms after feedback onset), consistent with the spectral, spatial and temporal locus of the positive feedback beta-gamma effect<sup>45,16</sup>. Within the beta-gamma band, the frequency with maximum difference in mean total power between reward and punishment trials was extracted on an individual participant basis and used as the target modulation frequency. Individually frequency-tuned modulation was applied with a resolution of 0.5 Hz, consistent with prior work<sup>10</sup>. All EEG processing and analysis were conducted using MATLAB calls on the FieldTrip toolbox<sup>60</sup>.

**Primary dependent measures.** In experiment 1, participants' choices for each trial were averaged across the three runs in each block, resulting in a learning curve composed of 40 choice scores for each participant, for each reward and punishment trial type, and for each pre-modulation, modulation and post-modulation time period. In experiment 2, OCI-R total score and subscales were generated by adding the item scores based on a five-point Likert scale.

For both experiments, we examined participants' beta-gamma effect, that is, the difference in the beta-gamma power (20–35 Hz) following reward feedback relative to punishment feedback (200–500 ms after feedback onset) at electrode Fz, during the pre-experiment period, and its relationship with reward optimal choice behavior and OCI-R total score.

**Computational modeling.** Four models are used; that is, two variants of a standard Q-learning model (StandardQ and WeightedQ)<sup>22,23</sup>, an Actor-Critic model<sup>22,23</sup>, and a Hybrid model incorporating both Q-learning and actor-critic elements<sup>23,24</sup> to best match the pattern of choice behavior in the reinforcement learning task.

**Q-learning models.** For every participant, the Q-learning algorithm<sup>22</sup> calculates the expected reward value ( $Q$ ) for an action ( $a$ ) based on their choice and feedback history. If the expected reward value for an action  $a$  in trial  $t$  is  $Q(a, t)$ , then the expected reward value for the same action in trial  $t + 1$  is computed as per the following rule:

$$Q(a, t + 1) = Q(a, t) + a_o \times \delta(t),$$

where  $a_o$  represents the learning rate and quantifies the degree to which feedback in trial  $t$  leads to adjustments in the expected value during trial  $t + 1$ . The value of  $a_o$  ranges between 0 and 1, with 0 indicating no influence and 1 indicating maximum influence.  $\delta(t)$  represents the prediction error computed as the difference between the actual outcome following the action  $a$  in trial  $t$  and the expected reward value as follows:

$$\delta(t) = \text{outcome}(t) - Q(a, t).$$

We used two variants of the Q-learning algorithm. In the StandardQ variant, the outcome( $t$ ) variable was set to 1 for positive feedback, 0 for neutral feedback, and  $-1$  for negative feedback. In the WeightedQ variant, we allowed positive and negative feedback to be weighted differently using an additional free parameter,  $d$ <sup>23</sup>. The outcome( $t$ ) value was set to  $1 - d$  for positive feedback, 0 for neutral feedback, and  $-d$  for negative feedback, allowing exclusive weighting for positive and negative outcomes when  $d = 0$  and  $d = 1$ , respectively, and equal weighting of positive and negative outcomes when  $d = 0.5$ . In both variants, the expected values ( $Q$ ) were set to 0 at the beginning of every run.

When choosing between two actions,  $a_1$  and  $a_2$ , in a given trial  $t$ , both models computed the expected reward value for each action,  $Q(a_1, t)$  and  $Q(a_2, t)$ , using the choice and feedback history as described above. Using these expected values, the probability of choosing the action  $a_1$  in that trial was computed using the following softmax rule:

$$P(a_1, t) = \exp(Q(a_1, t)/\beta) / [\exp(Q(a_1, t)/\beta) + \exp(Q(a_2, t)/\beta)].$$

Here, the parameter  $\beta$  represents the noise or the temperature during the choice process. Higher  $\beta$  values suggest greater randomness in the choice process and lesser sensitivity to expected reward values. Lower  $\beta$  values suggest higher propensity to choose the stimulus with larger expected rewards.

**Actor-Critic model.** In the Actor-Critic model, every pair of stimuli is represented as a state  $s$ , and the expected value associated with that state in a given trial  $t$  is represented as  $V(s, t)$ . The value functions are updated using a prediction error,  $\epsilon(t)$ , as follows:

$$V(s, t + 1) = V(s, t) + a_c \times \epsilon(t),$$

$$\epsilon(t) = \text{outcome}(t) - V(s, t),$$

where  $a_c$  represents the learning rate of the critic module, which determines the degree to which state values are updated according to the prediction error, calculated as the difference between the expected value for the current state,  $V(s, t)$ , and the actual outcome, outcome( $t$ ). Similar to the WeightedQ model, outcome( $t$ ) was set to  $1 - d$  for positive feedback, 0 for neutral feedback, and  $-d$  for negative feedback.

The actor module selects from among multiple actions,  $a$ , within a state,  $s$ , on a trial,  $t$ , according to their weights,  $w(s, a, t)$ . The weight of the selected action is adjusted using the prediction error from the critic module,  $\epsilon(t)$ , using the following rule:

$$w(s, a, t + 1) = w(s, a, t) + a_A \times \epsilon(t),$$

where  $a_A$  is the learning rate for the actor module, which is used to determine the degree of influence of the prediction error on the action weight. Following ref. <sup>23</sup>, we normalized the actor weights using the sum of their absolute values to avoid unbounded growth, and the weights were initialized at 0.01. For example, given two possible actions,  $a_1$  and  $a_2$ , which reflect selection of stimulus 1 or stimulus 2, respectively, actor weight for action 1 was normalized as follows:

$$w(s, a_1, t) \leftarrow w(s, a_1, t) / (|w(s, a_1, t)| + |w(s, a_2, t)|).$$

Using these weights, the probability of choosing the action  $a_i$  in a trial  $t$ ,  $P(a_i, t)$ , was determined using the following softmax rule:

$$P(a_i, t) = \exp(w(s, a_i, t)/\beta) / [\exp(w(s, a_1, t)/\beta) + \exp(w(s, a_2, t)/\beta)],$$

where the  $\beta$  parameter reflects the temperature or the degree of exploration during the choice.

**Hybrid actor-critic Q-learning model.** We used a Hybrid model<sup>23,24</sup>, which incorporates the expected value of each action from the WeightedQ model with the action weights determined from the Actor-Critic model, therefore this reflects an interaction between model-based and model-free learning mechanisms. The expected values of each action in the WeightedQ model were combined with the weights of each action in the Actor-Critic model using a mixing parameter,  $c$ , the value of which reflects the independent contributions of each model as per the following rule:

$$H(s, a, t) = [(1 - c) \times w(s, a, t) + c \times Q(a, t)].$$

As  $c$  ranges from 0 to 1, the model transitions from a purely Actor-Critic model to the WeightedQ model. Action  $a_i$  is then selected in a given trial  $t$  with a probability  $P(a_i, t)$  using a softmax function:

$$P(a_i, t) = \exp(H(s, a_i, t)/\beta) / [\exp(H(s, a_1, t)/\beta) + \exp(H(s, a_2, t)/\beta)],$$

where the  $\beta$  parameter reflects the temperature or the degree of exploration during the choice.

**Parameter estimation.** Using maximum likelihood estimation with the MATLAB function *fmincon*, we optimized two parameters in the StandardQ model ( $a, \beta$ ), three parameters in the WeightedQ model ( $a_\alpha, \beta, d$ ), four parameters in the Actor-Critic model ( $a_c, a_\alpha, \beta, d$ ), and six parameters in the Hybrid model ( $a_\alpha, a_c, a_\alpha, \beta, d, c$ ). The parameters were estimated such that the probability of actual choices was maximized for reward trials. These estimations were performed for each run in every block (pre-modulation, modulation, post-modulation) for participants in every group (sham, alpha, beta-gamma). We also performed model fits after combining reward and punishment trials and the pattern of results remained similar.

**Statistics.** In experiment 1, we carried out repeated measures ANOVA using the within-participants factors of time (pre-modulation versus modulation versus post-modulation) and valence (reward versus punishment), and the between-participants factor of modulation group (passive sham versus active control/alpha versus personalized beta-gamma) to assess optimal choice performance. The analysis of temperature used the within-participants factor of time (pre-modulation versus modulation versus post-modulation) and the between-participants factor of modulation group (passive sham versus active control/alpha versus personalized beta-gamma). When appropriate, follow-up ANOVAs were conducted to test specific preplanned hypotheses. To examine the relationship between reward-guided optimal choice behavior and the reward feedback beta-gamma effect, we conducted a participant-wise two-tailed Pearson correlation between the mean EEG beta-gamma power in the pre-experiment reward feedback (20–35 Hz, 200–500 ms after feedback onset) and optimal choice accuracy on reward trials during the pre-modulation period, collapsed across modulation groups (passive sham, active control/alpha, personalized beta-gamma). Participant-wise two-tailed Pearson correlation was used to examine the association between reward optimal choice behavior during the pre-modulation and the modulation periods for the beta-gamma group.

In experiment 2, we carried out repeated measures ANOVA for the OCI-R total and subscale scores using the within-participants factor of time (pre-modulation versus day 5 versus 1 month versus 2 months versus 3 months) and the between-participants factor of modulation group (active control/alpha versus personalized beta-gamma). Interactions were parsed with Bonferroni-corrected pairwise comparisons. Participant-wise two-tailed Pearson correlations, using an adjusted  $P$  value ( $P = 0.0008$ ) to control for multiple comparisons, were conducted to examine relationships between participants' pre-modulation baseline OCI-R scores and each post-modulation OCI-R score. To examine the relationship between baseline OCI-R total scores and pre-experiment reward choice behavior, the data were collapsed across modulation groups (active control/alpha, personalized beta-gamma) and an ANOVA was run on pre-experiment optimal choice behavior from reward trials using the between-participants factor of OCI-R score (low versus high). Participant-wise Pearson correlations (two-tailed) were conducted to examine the association between individual participant pre-experiment reward choice behavior and baseline OCI-R total score. To compare the EEG beta-gamma effect (20–35 Hz, 200–500 ms after feedback onset) between individuals with high and low baseline OCI-R total scores, we conducted an ANOVA using the within-participants factor of valence (reward versus punishment) and the between-participants factor of OCI-R score (low versus high). For both experiments, ANOVA  $P$  values were adjusted using the Greenhouse-Geisser

epsilon correction for non-sphericity when this assumption was violated. The experiments of this study were not repeated independently. However, partial eta squared effect sizes were estimated to facilitate comparison between studies and promote replication. All tests were two-tailed with the exception of  $F$ -tests, which by definition are one-tailed. Statistical parameters, including the definitions and exact value of  $n$  (that is, total number of participants),  $P$  values, and the type of statistical tests, are reported primarily in the figure legends.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The data used for analysis in this study are freely and permanently available on Open Science Framework (<https://osf.io/ctph2/>). Source data are provided with this paper.

## Code availability

The software code used in this study is freely and permanently available on Open Science Framework (<https://osf.io/ctph2/>).

## References

- Insel, T. et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
- Toffolo, M. B. J., van den Hout, M. A., Engelhard, I. M., Hooge, I. T. C. & Cath, D. C. Uncertainty, checking, and intolerance of uncertainty in subclinical obsessive compulsive disorder: an extended replication. *J. Obsessive Compuls. Relat. Disord.* **3**, 338–344 (2014).
- Frane, J. W. A method of biased coin randomization, its implementation, and its validation. *Ther. Innov. Regul. Sci.* **32**, 423–432 (1998).
- Karabanov, A. N., Saturnino, G. B., Thielscher, A. & Siebner, H. R. Can transcranial electrical stimulation localize brain function? *Front. Psychol.* **10**, 213 (2019).
- Dmochowski, J. P., Datta, A., Bikson, M., Su, Y. & Parra, L. C. Optimized multi-electrode stimulation increases focality and intensity at target. *J. Neural Eng.* **8**, 046011 (2011).
- Edwards, D. et al. Physiological and modeling evidence for focal transcranial electrical brain stimulation in humans: a basis for high-definition tDCS. *Neuroimage* **74**, 266–275 (2013).
- Villamar, M. F. et al. Focal modulation of the primary motor cortex in fibromyalgia using 4×1-ring high-definition transcranial direct current stimulation (HD-tDCS): immediate and delayed analgesic effects of cathodal and anodal stimulation. *J. Pain* **14**, 371–383 (2013).
- Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
- Huppert, J. D. et al. The OCI-R: validation of the subscales in a clinical sample. *J. Anxiety Disord.* **21**, 394–406 (2007).
- Chasson, G. S., Tang, S., Gray, B., Sun, H. & Wang, J. Further validation of a Chinese version of the obsessive-compulsive inventory-revised. *Behav. Cogn. Psychother.* **41**, 249–254 (2013).
- Reinhart, R. M. G. Disruption and rescue of interareal theta phase coupling and adaptive behavior. *Proc. Natl Acad. Sci. USA.* **114**, 11542–11547 (2017).
- Reinhart, R. M. G., Cosman, J. D., Fukuda, K. & Woodman, G. F. Using transcranial direct-current stimulation (tDCS) to understand cognitive processing. *Atten. Percept. Psychophys.* **79**, 3–23 (2017).
- Reinhart, R. M. G., Xiao, W., McClenahan, L. & Woodman, G. F. Electrical stimulation of visual cortex can immediately improve spatial vision. *Curr. Biol.* **25**, 1867–1872 (2016).
- Dayan, E., Censor, N., Buch, E. R., Sandrini, M. & Cohen, L. G. Noninvasive brain stimulation: from physiology to network dynamics and back. *Nat. Neurosci.* **16**, 838–844 (2013).
- Poreisz, C., Boros, K., Antal, A. & Paulus, W. Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. *Brain Res. Bull.* **72**, 208–214 (2007).
- Gandiga, P. C., Hummel, F. C. & Cohen, L. G. Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clin. Neurophysiol.* **117**, 845–850 (2006).
- Ali, M. M., Sellers, K. K. & Fröhlich, F. Transcranial alternating current stimulation modulates large-scale cortical network activity by network resonance. *J. Neurosci.* **33**, 11262–11275 (2013).
- Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
- Kayser, J. & Tenke, C. E. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clin. Neurophysiol.* **117**, 348–368 (2006).

59. Srinivasan, R., Winter, W. R., Ding, J. & Nunez, P. L. EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. *J. Neurosci. Methods* **166**, 41–52 (2007).
60. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 1–9 (2011).

### Acknowledgements

This work was supported by a grant from the National Institutes of Health (R01-MH114877) and a generous gift from an individual philanthropist awarded to R.M.G.R.

### Author contributions

R.M.G.R. and S.G. conceived the experiments and analyses. R.M.G.R., J.A.N. and V.V. performed the experiments. R.M.G.R., S.G. and J.A.N. analyzed and interpreted the data, and R.M.G.R. and S.G. wrote the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

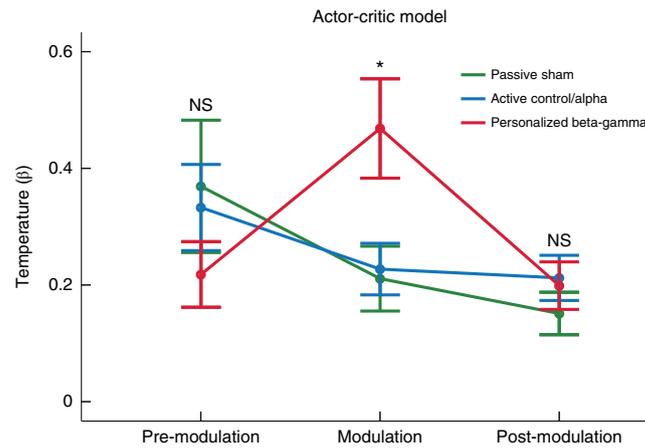
**Extended data** is available for this paper at <https://doi.org/10.1038/s41591-020-01173-w>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41591-020-01173-w>.

**Correspondence and requests for materials** should be addressed to R.M.G.R.

**Peer review information** *Nature Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. Kate Gao and Jerome Staal were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | Neuromodulation effects on temperature.** Mean temperature shown across pre-modulation, modulation, and post-modulation periods for each group (passive sham, green; active control/alpha, blue; personalized beta-gamma, red). A significant time  $\times$  group interaction was observed ( $F_{4,114} = 3.066$ ,  $P = 0.027$ ,  $\eta_p^2 = 0.097$ ,  $n = 60$ ). There were significant time  $\times$  group interactions for the beta-gamma and alpha groups ( $F_{2,76} = 4.942$ ,  $P = 0.012$ ,  $\eta_p^2 = 0.115$ ,  $n = 40$ ) and the beta-gamma and sham groups ( $F_{2,76} = 4.119$ ,  $P = 0.027$ ,  $\eta_p^2 = 0.098$ ,  $n = 40$ ), but not for the alpha and sham groups ( $F_{2,76} = 0.282$ ,  $P = 0.686$ ,  $\eta_p^2 = 0.007$ ,  $n = 40$ ). A significant enhancement in temperature was evident only during beta-gamma HD-tACS relative to sham ( $F_{1,38} = 6.409$ ,  $P = 0.016$ ,  $\eta_p^2 = 0.144$ ,  $n = 40$ ) or to alpha ( $F_{1,38} = 6.311$ ,  $P = 0.016$ ,  $\eta_p^2 = 0.142$ ,  $n = 40$ ). The effect was rapidly extinguished upon switching off HD-tACS, given that no differences were observed between the alpha and beta-gamma ( $F_{1,38} = 0.056$ ,  $P = 0.814$ ,  $\eta_p^2 = 0.001$ ,  $n = 40$ ) or the sham and beta-gamma groups ( $F_{1,38} = 0.762$ ,  $P = 0.388$ ,  $\eta_p^2 = 0.02$ ,  $n = 40$ ) in the post-modulation period. Of note, baseline temperature values were relatively stable and did not significantly differ between groups during the pre-modulation period (alpha versus beta-gamma,  $F_{1,38} = 1.528$ ,  $P = 0.224$ ,  $\eta_p^2 = 0.039$ ,  $n = 40$ ; alpha versus sham,  $F_{1,38} = 0.072$ ,  $P = 0.790$ ,  $\eta_p^2 = 0.002$ ,  $n = 40$ ; sham versus beta-gamma,  $F_{1,38} = 1.421$ ,  $P = 0.241$ ,  $\eta_p^2 = 0.036$ ,  $n = 40$ ). No other parameters showed significant effects. Mixed ANOVAs used the within-participants factor of time (pre-modulation, modulation, post-modulation) and the between-participants factor of group (sham, alpha, beta-gamma). Follow-up univariate ANOVAs within individual modulation periods used the between-participants factor of group (alpha, beta-gamma; alpha, sham; or beta-gamma, sham). Error bars show  $\pm 1$  s.e.m. \* $P < 0.05$ . NS, not significant.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data were collected with readily available software (Psychtoolbox-3, Brainard, 1997, RRID SCR\_002881; MATLAB 2018a, Mathworks, RRID SCR\_001622).

Data analysis

Data were analyzed with readily available software (FieldTrip\_20200215, Oostenveld et al., 2011, <http://www.fieldtriptoolbox.org/>; HD-Explore and HD-Targets, Soterix Medical, New York, NY, USA, <https://soterixmedical.com/research/software>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data used for analysis in this study are freely available on Open Science Framework (<https://osf.io/ctph2/>). Additionally, statistical source data are provided for Figure 3, Figure 4, and Extended Data Figure 1.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications (Reinhart and Nguyen, 2019 Nat Neurosci; Violante et al. 2017 Elife; Helfrich et al. 2014 Public Library of Sciences, Biology; Polania et al. 2012 Current Biology).
Data exclusions	No data was excluded in Experiment 1. In Experiment 2, 68 participants were recruited such that 34 participants had low OCI-R scores, and the remaining 34 had high OCI-R scores. Of the 34 participants with low OCI-R scores, four participants were lost to attrition due to multiple testing sessions. It is common in multi-day human neuroscience studies that not all subjects will be retained. All inclusion/exclusion criteria used in the study were pre-established.
Replication	Experiment 1 and Experiment 2 were not directly and independently replicated within this study using an additional cohort of participants, which is common practice in multi-day, human noninvasive brain stimulation and cognitive neuroscience research. However, indirectly, we did successfully determine the critical findings linking reward choice behavior and beta-gamma neural feedback effects across experiments as well as the findings linking reward choice behavior and the effects of beta-gamma neuromodulation across experiments. Further, in Experiment 2, the primary outcome measures (total and subscale OCI-R scores) were tracked for up to 3 months following the intervention to confirm the stability and reliability of observed effects. Pearson two-tailed correlations between all time intervals in the alpha control group showed good to excellent reliability for total and subscale scores (0.526 to 0.947, $P_s < 0.002$ ), consistent with previous test-retest reliability studies of OCI-R in clinical and subclinical OCD populations (Foa et al. 2002; Hajcak et al. 2004).
Randomization	Participants were randomly assigned to experimental groups using a covariate adaptive randomization method (Frane, 1998) to produce equal-sized groups balanced by sex (Experiments 1 and 2), age (Experiments 1 and 2), and OCI-R score (Experiment 2). Reward and punishment trials were randomly interleaved within a run during the reinforcement learning task. Duration of the intertrial interval was randomly jittered with a rectangular distribution in the reinforcement learning task.
Blinding	The experiments were double-blind (both subject and experimenter blinding was instituted). Both experiments were sham-controlled with Experiment 1 consisting of both passive and active sham procedures, and Experiment 2 consisting of an active sham procedure. The active sham procedure effectively minimizes unwanted effects associated with changes in extra-cranial tissue, shunting, and peripheral co-stimulation, such as transretinal or transcutaneous stimulation (Karabanov et al., 2019). Both experiments followed a between-participants design but also included within-participant pre- and post-intervention measurements of the primary outcome.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

### Methods

- n/a | Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants
- Clinical data

- n/a | Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Experiment 1 consisted of 60 participants (sham, $n = 20$ , mean age $26.6 \pm 5.8$ s.d., 10 female, all right-handed; alpha, $n = 20$ , mean age $25.8 \pm 5.6$ s.d., 10 female, all right-handed; beta-gamma, $n = 20$ , mean age $25.1 \pm 6.2$ s.d., 11 female, all right-handed).  Experiment 2 consisted of 68 participants with 34 participants were recruited with relatively low OCI-R scores (0-17) and another 34 participants with relatively high OCI-R scores (18 or higher). Four participants were lost to attrition due to multiple testing sessions. The final sample of 64 participants consisted of 30 people with low scores (active control, $n = 15$ , mean age $23.9 \pm 4.0$
----------------------------	---

s.d., 8 female, all right-handed; beta-gamma,  $n = 15$ ,  $25.3 \pm 3.8$  s.d., 8 female, all right-handed), and 34 people with high scores (active control,  $n = 17$ ,  $23.4 \pm 4.0$  s.d., 9 female, all right-handed; beta-gamma,  $n = 17$ ,  $23.0 \pm 3.7$  s.d., 9 female, all right-handed).

#### Recruitment

Subjects were recruited from the greater Boston area via posting announcements on local and electronic bulletin boards and were selected based on the following eligibility requirements: no metal implants in head, no implanted electronic devices, no history of neurological problems or head injury, no skin sensitivity, no claustrophobia, not being pregnant, no history of an Axis I psychiatric disorder, no current use of psychoactive medication, no substance abuse or dependence within the last 6 months, normal color vision, normal or corrected-to-normal visual acuity, and not being enrolled in any other research or treatment study during the duration of their study participation. As with all human research requiring volunteers, the possibility of introducing self-selection bias is unavoidable. As a result, it is possible that individuals inclined to sensation-seeking or thrill-seeking are more likely to pursue research studies requiring volunteers. However, in the present study, these personality traits were not explicitly under study.

#### Ethics oversight

Boston University Institutional Review Board, Charles River Campus.

Note that full information on the approval of the study protocol must also be provided in the manuscript.