Trace Imbalance in Reinforcement and Punishment Systems Can Mis-reinforce Implicit Choices Leading to Anxiety

Yuki Sakai[a,b,†], Yutaka Sakai[c,†], Yoshinari Abe[b], Jin Narumoto[b], Saori C. Tanaka[a,*]

[a]ATR Brain Information Communication Research Laboratory Group, 2-2-2 Hikaridai Seika-Cho, Soraku-Gun, Kyoto 619-0288, Japan

[b]Department of Psychiatry, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, 465 Kajii-Cho, Kawaramachi-Hirokoji, Kamigyo-Ku, Kyoto 602-8566, Japan

[c]Brain Science Institute, Tamagawa University, 6-1-1, Tamagawa-gakuen, Machida, Tokyo 194-8610, Japan

[†]These authors contributed equally and should be considered co-first authors.

*Corresponding author: Saori C. Tanaka

Computational Neuroscience Laboratories, Advanced Telecommunications Research Institute International, 2-2-2 Hikaridai, Seika-Cho, Soraku-Gun, Kyoto 619-0288, Japan.

Tel: +81-774-95-1250; Fax: +81-774-95-1236.

E-mail: xsaori@atr.jp

Abstract

Nobody wants to experience anxiety. However, anxiety may be induced by our own implicit choices that are mis-reinforced by some imbalance in reinforcement learning. Here we focused on obsessive-compulsive disorder (OCD) as a candidate for implicitly learned anxiety. Simulations in the reinforcement learning framework showed that agents implicitly learn to become anxious when the memory trace signal for past actions decays differently for positive and negative prediction errors. In empirical data, we confirmed that OCD patients showed extremely imbalanced traces, which were normalized by serotonin enhancers. We also used fMRI to identify the neural signature of OCD and healthy participants with imbalanced traces. Beyond the spectrum of clinical phenotypes, these behavioral and neural characteristics can be generalized to variations in the healthy population.

Introduction

Humans sometimes experience anxiety. Given that no one likes to feel anxious, we tend to believe that anxiety is passively driven by external factors. However, is it possible that our own choices implicitly induce anxiety? We make mental choices when thinking or mind-wandering, as well as in response to the external world. We are always learning appropriate choices through our experiences, and some choices are learned in subliminal situations[1, 2]. Indeed, some evidence suggests that many of our decisions and actions in everyday life are shaped by implicit learning processes[3]. Hence, our anxiety might be induced by our own choices that are reinforced through implicit learning.

When investigating the possibility of implicit choices leading to anxiety, we must consider its theoretical and biological implementation. Implicit learning requires repetitive experiences[1, 2]. Reinforcement learning theory provides a framework for learning appropriate choices through repetitive experiences[4]. Reinforcement learning reinforces or punishes recent choices based on the difference in the actual outcome from the prediction, called the prediction error. A positive prediction error reinforces recent choices, whereas a negative error punishes them. When examining the biological implementation of implicit learning, it is essential to consider previous work demonstrating that the activity of dopamine-projecting neurons in the midbrain resembles prediction error[5]. Dopaminergic neurons widely project to cortical and subcortical areas. One of the main targets of dopamine is the striatum, which is located in the basal ganglia[6]. In the striatum, distinct types of dopamine receptors—D1 and D2 receptors—are expressed in exclusive neuron groups, which are considered to be involved in distinct circuit patterns, namely, direct and indirect pathways[6]. These distinct neural systems are thought to play different roles in reinforcement learning. Artificial activations of D1- and D2-expressing neurons after a certain behavioral action reinforce and punish the action, respectively[7]. Direct and indirect pathway neurons respond to positive and negative outcomes, respectively[8]. Dopamine-dependent synaptic plasticity in corticostriatal synapses, which underlies reinforcement learning, shows opposing dopamine dependence for D1 and D2 neurons[9, 10]. These results imply that reinforcement and punishment of recent actions may be reflected through different neural systems: the direct and indirect pathways.

In the implementation of implicit learning, we should also consider delay as well as reinforcement/punishment. In general, the outcome of a certain choice is available after a certain delay. Therefore, reinforcement and punishment should be assigned to recent choices within a certain time scale. This is called credit assignment, which is implemented as eligibility traces in reinforcement learning[4]. An eligibility trace can be implemented as a memory trace in each synapse[10-12]. In this case, the trace time scales for reinforcement and punishment should be separately controlled in direct and indirect pathways. Reinforcement learning theory requires that both trace time scales be equal. However, the requirement cannot be completely realized in distinct neural systems.

Here, we propose a computational model that enables the implementation of imbalanced learning, and we aimed to determine whether the imbalance induces some defects in specific situations. We focused on obsessive-compulsive disorder (OCD) as a candidate for such an imbalanced condition for two reasons. First, abundant evidence suggests that imbalance between direct and indirect pathways is central to the pathophysiology of OCD[13-15]. Second, obsession in OCD can be an implicit choice leading to anxiety. In patients with OCD, an intrusive thought becomes lodged in their mind and drives obsessive anxiety[16]. To neutralize and relieve the anxiety, they tend to perform a compulsive action, despite its cost (e.g., excessively washing their hands or repeatedly checking their keys), even though the anxiety will spontaneously diminish before long if no action is taken[17].

This article proposes a computational model to reproduce the abnormal repetition of obsession and compulsion in OCD symptoms, incorporating two hypotheses: (1) obsession is induced by patients' implicit choices, and (2) eligibility trace time scales are imbalanced for reinforcement and punishment in the reinforcement learning framework. We demonstrate that the imbalance in trace time scales mis-reinforces implicit choices leading to obsession, resulting in a spiral of repetitive obsession and compulsion. We tested our hypothesis in a behavioral task for healthy participants and participants with OCD and evaluated the neural substrates of imbalanced eligibility trace time scales using resting-state functional magnetic resonance imaging (rs-fMRI).

Results

OCD-like behavior in a separate eligibility trace model

Eligibility traces determine the credit assignment of the outcome prediction error of recent actions[4]. Each trace represents the recent frequency of an action in a specific state within a time scale determined by trace decay factor ν (see Online Methods). When a prediction error occurs, the choice probability of each action in each state is updated by the product of the eligibility trace and the prediction error (Figure 1a). Thus, a positive prediction error reinforces recent actions within a time scale, whereas a negative error punishes them. We assumed that the eligibility traces for reinforcement and punishment might be implemented in distinct neural systems (red and blue in Figure 1a; see Online Methods). Theoretically, the trace factors in the two systems, $\nu^{\pm}$, should be balanced: $\nu^{+} = \nu^{-}$. However, the separate neural implementation makes it difficult to perfectly maintain the balance. Here, we assumed that the balance was not maintained: $\nu^{+} \neq \nu^{-}$.

We modeled mental states involved in anxiety into stochastic transitions between two states: relief and anxiety (Figure 1b). We assumed an action to relieve anxiety as an option in the anxiety state, which could stochastically permit a transition from the anxiety state to the relief state at a certain cost. The action might become compulsive in OCD, and we labeled it "compulsion". We unified any other options in the anxiety state which could relieve the anxiety at lower probability without any cost into the option "other". Our fundamental assumption was that the transition from the relief state to the anxiety state would be caused by the individual’s own choice, which might become obsessive in OCD. We labeled this option "obsession". We also unified any other options in the relief state which maintained the relief state into "other". We assumed that every stay in the anxiety state would produce a negative outcome.

No positive outcome was introduced in the anxiety-relief transition model. Hence, maintaining "other" in the relief state is clearly optimal. Normal reinforcement learning should reduce the obsession rate. However, if the learning system has some defects, such as an imbalance in trace factors, then the learning system may fail to reduce the obsession and fall into a spiral of obsession and compulsion. To show this possibility, we simulated actor-critic learning (see Online Methods), a typical method of reinforcement learning models[4] in the anxiety-relief state-transition (Figure 1b), incorporating a separate eligibility trace (Figure 1a). When the imbalance in the trace factors was moderate, actor-critic learning was able to reduce the obsession rate, even when the simulation started at a high obsession rate (Figure 1c). In contrast, extreme imbalance induced an increase in the obsession rate that culminated in a spiral of repetitive obsession and compulsion (Figure 1d). Such OCD-like behavior was reduced by preventing the compulsion during exposure to the anxiety state (green bar in Figure 1d). This is the behavioral therapy of exposure and response prevention (ERP) and is one of the first-line treatments of OCD[18].

We identified the condition for the trace factors $\nu^{\pm}$ in which OCD-like behavior might emerge by numerical simulations in a representative set of other model parameters (Figure 1e; see Online Methods). In addition to actor-critic learning (leftmost in Figure 1e), we simulated SARSA (State-Action-Reward-State-Action) and Q-learning (right in Figure 1e), other typical reinforcement learning models[4]. The color map represents the fraction of 100 simulation runs in which obsession was reinforced at each pair of (ν^{+}, ν^{-}). We identified the conditions of OCD-like behavior in the region $\nu^{+} > \nu^{-}$ (Figure 1e). Namely, the trace scale for reinforcement was longer than that for punishment. We also identified the condition in which the behavioral therapy of ERP would not work and localized it to a region with more robust imbalance (Figure 1f).

We obtained the theoretical conditions as functions of the other parameters (see the Supplementary Note). Each OCD patient presents a certain compulsive action for specific anxieties among many types of anxiety. Therefore, the corresponding parameters of the anxiety-relief transition model (Figure 1b) should differ by the type of anxiety. Next, we derived the condition for the learning parameters in which OCD-like behavior emerges in some type of anxiety in actor-critic, SARSA, and Q-learning and obtained the common condition that $\nu^{+} > \nu^{-}$ (see Section 3.3 in the Supplementary Note). This result suggests that a person with $\nu^{+} > \nu^{-}$ has a risk of OCD.

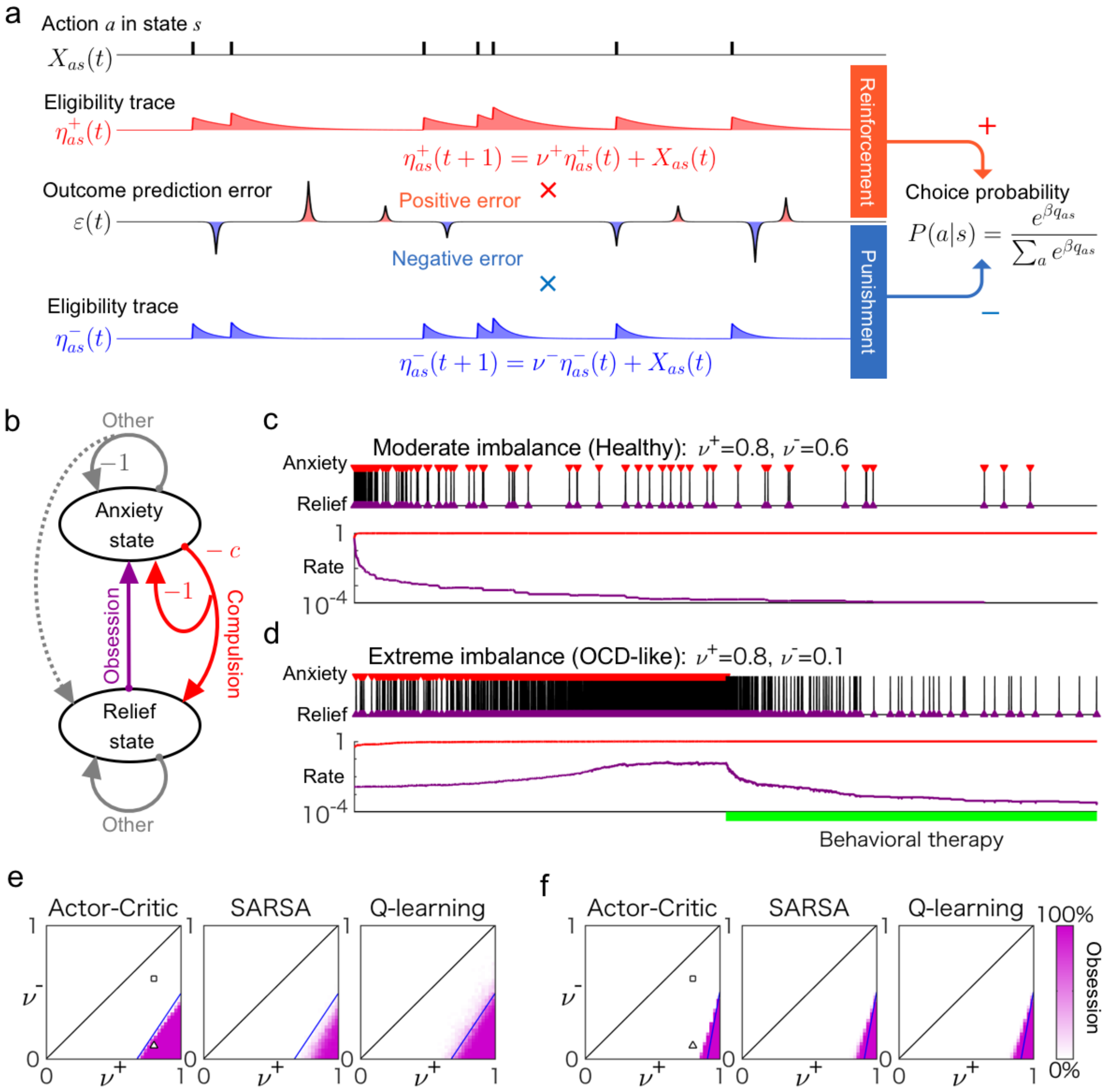

Figure 1. OCD-like behavior in a separate eligibility trace model. (a) Schema of the separate eligibility trace model. (b) State transitions in the anxiety-relief model. (c) Simulation of the separate eligibility trace model in the anxiety-relief state-transition in the case of a moderate imbalance in the trace factors $\nu^{\pm}$. The black polylines represent the temporal pattern of state transitions. The red and magenta triangles represent the occurrences of compulsions and obsessions, respectively. The lower plots show the temporal patterns of the compulsion (red

curve) and obsession (magenta curve) rates on a logarithmic scale. (d) As in (c), but in the case of extreme imbalance. In the latter half, we demonstrated behavioral therapy by preventing the compulsion in the anxiety state (green bar). (e) Conditions of OCD-like behavior in the space of the trace factors $\nu^{\pm}$ with different types of learning algorithms: actor-critic, SARSA, and Q-learning. The magenta color scale represents the percentage of the obsession rate increase among 100 times simulations around optimal behavior (zero obsession rate). Solid blue lines represent the theoretically derived boundary (see the Supplementary Note). The square and triangle represent the trace factors $\nu^{\pm}$ used in (c) and (d). (f) As in (e), but, in this condition, the behavioral therapy did not work.

Participants and behavioral task

To test the suggestion that $\nu^{+} > \nu^{-}$, derived from our computational model of OCD, we applied the delayed feedback task to patients with OCD (n = 33) and healthy controls (HCs) (n = 168). Fifteen HCs were excluded for medical or experimental reasons, and the subsequent analysis was conducted in 33 OCD patients and 153 HCs (Supplementary Table 1). Considering the therapeutic effect of the serotonin reuptake inhibitor (SRI) in OCD[18], SRI might normalize the imbalanced setting of $\nu^{+} > \nu^{-}$. Because a meta-analysis of SRI treatment suggested that higher doses of SRI are more effective in the treatment of OCD versus other psychiatric conditions, such as major depressive disorder[19], we divided the OCD patients into two groups by the dose equivalence of SRIs[20] (Online Methods and Supplementary Table 2): OCD patients with higher SRI doses were grouped into $\mathrm{OCD}_{\mathrm{HighSRI}}$ (n = 10), and those with lower SRI doses or no psychotropic medications were $\mathrm{OCD}_{\mathrm{Low\text{-}NoSRI}}$ ($\mathrm{OCD}_{\mathrm{LowSRI}}$, n = 10; $\mathrm{OCD}_{\mathrm{NoSRI}}$, n = 13). There were no significant differences in obsessive-compulsive and depressive symptoms evaluated by the Yale-Brown Obsessive-Compulsive Scale[21] and the 17-item Hamilton Depression Rating Scale (HDRS)[22] between the $\mathrm{OCD}_{\mathrm{HighSRI}}$ and $\mathrm{OCD}_{\mathrm{Low\text{-}NoSRI}}$ groups (Supplementary Table 1). In addition, no patients had current major depressive disorder as a comorbidity (Online Methods).

We used the delayed feedback task to evaluate the trace factors $\nu^{\pm}$ as in our previous research[23] except for the presented stimuli (abstract cues). Briefly, participants chose one of two options displayed on the screen by pressing a left or right button in each trial (Figure 2a). Each stimulus represented different outcomes (+10, +40, −10, or −40 yen) and a delay to the feedback (immediately after a button press or three trials later) (Figure 2b and c). For example, +40(0) represented a gain of 40 yen within the current trial (immediate reward) and −10(3) represented a loss of 10 yen after three trials (delayed punishment). The participants were not told about the stimulus-outcome associations shown in Figure 2b. They received money after the experiment in proportion to the total outcome obtained in the delayed feedback task. To maximize the total outcome, the participants needed to learn the appropriate stimulus choices by correctly assigning the credit of the current feedback to the recent choices that caused the current feedback. The difference in learning effects between immediate and delayed feedbacks reflected the trace factors $\nu^{\pm}$.

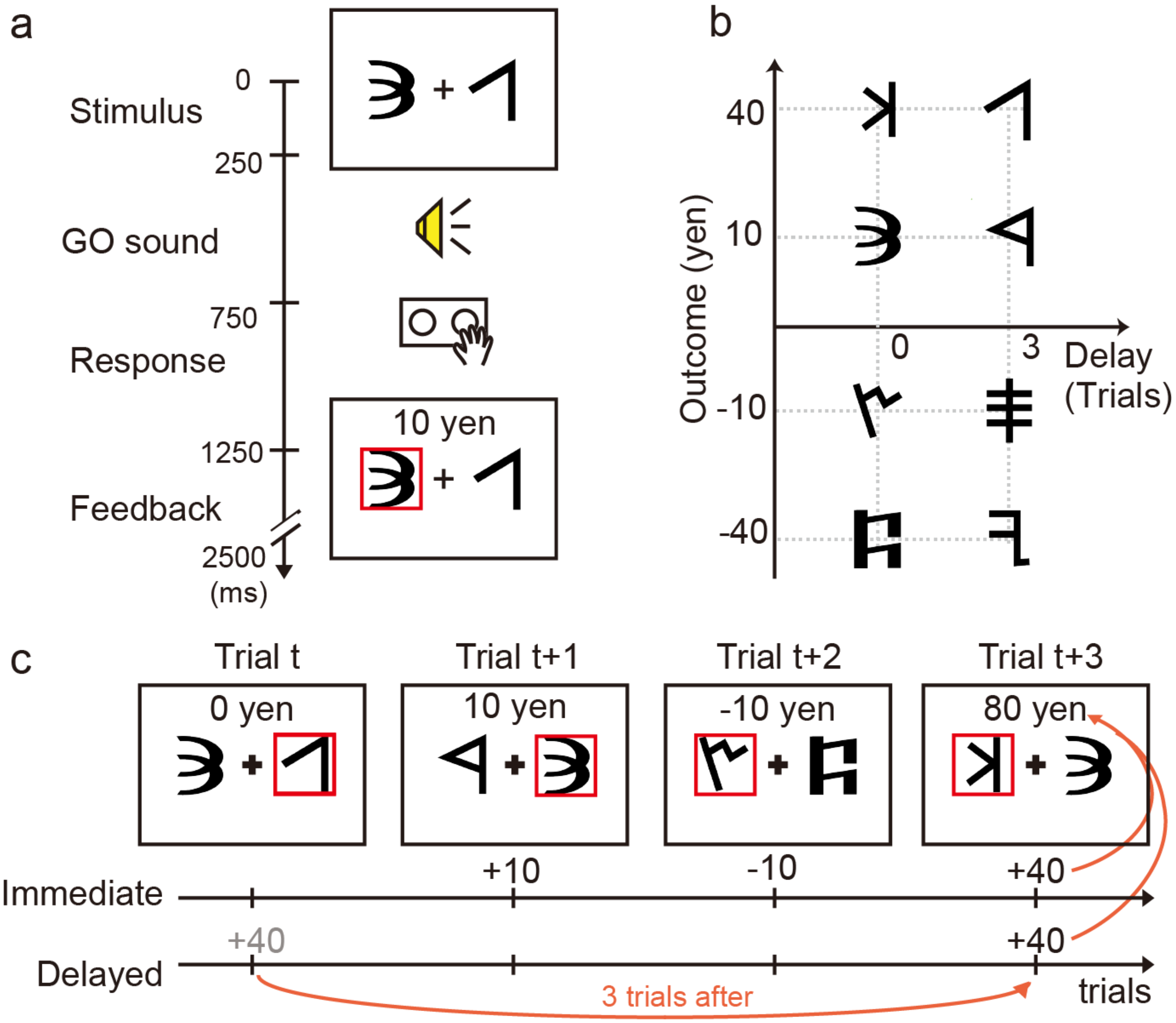

Figure 2. Delayed feedback task. (a) Two abstract cues were displayed in each trial. When the participants heard an auditory cue (beep), they chose one of the stimuli within 1 s. A single trial lasted 2.5 s. The sequence of the stimuli pair was pseudorandom. (b) Outcome-delay mapping for each stimulus. The mapping was different among participants. (c) An example of a delayed feedback task. If participants chose the stimuli with a delay [+40(3)] at trial *t*, that outcome was not displayed immediately. If they chose the stimuli with no delay [+40(0)] at trial *t*+3, the sum of the delayed and immediate outcomes was displayed at trial *t*+3 (in this case +80 yen).

Behavioral results

Based on our simulation of OCD-like behavior with trace factors $\nu^{+} > \nu^{-}$ (Figure 1e), patients with OCD would show impaired learning in the stimuli with delayed feedback. Therefore, the optimal choice rates between the representative four pairs of stimuli with the same delay and different magnitudes [pairs with no delays: +40(0) vs. +10(0) and −10(0) vs. −40(0); pairs with delays: +40(3) vs. +10(3) and −10(3) vs. −40(3)] were compared among groups. Consistent with our trace factor hypothesis, the $\text{OCD}_{\text{Low-NoSRI}}$ group showed impaired learning of stimuli with delays and performed at chance levels (Figure 3). Notably, $\text{OCD}_{\text{HighSRI}}$ patients exhibited intact learning similar to HCs in all pairs of interests.

A mixed-design two-way repeated-measures ANOVA with a within-participant factor of sessions (1–6 sessions) and a between-participant factor of groups ($\text{OCD}_{\text{Low-NoSRI}}$ patients, $\text{OCD}_{\text{HighSRI}}$ patients, and HCs) was conducted to clarify the between-group differences. There were significant interactions in the learning of stimulus pairs with delays. In the 10(3) vs. 40(3) pair, the interaction between session and group ($F(7.05, 645.34) = 2.32$, $p = 0.024$, $\eta_p^2 = 0.025$) and the simple main effects of group were significant in sessions 4 ($F(2, 183) = 3.97$, $p = 0.021$, $\eta_p^2 = 0.042$) and 5 ($F(2, 183) = 6.27$, $p = 0.0023$, $\eta_p^2 = 0.064$). Bonferroni-Holm–corrected post-hoc comparisons confirmed $\text{OCD}_{\text{Low-NoSRI}}$ patients < HCs ($t(183) = 2.81$, $p_{\text{adjusted}} = 0.016$) in session 4 and $\text{OCD}_{\text{Low-NoSRI}}$ patients < HCs ($t(183) = 3.54$, $p_{\text{adjusted}} = 0.0015$) in session 5 (Figure 3).

Regarding the −10(3) vs. −40(3) pair, the interaction between session and group ($F(7.42, 678.8) = 2.09$, $p = 0.039$, $\eta_p^2 = 0.022$) and the simple main effects of group were significant in sessions 5 ($F(2, 183) = 7.59$, $p = 0.0007$, $\eta_p^2 = 0.077$) and 6 ($F(2, 183) = 5.83$, $p = 0.0035$, $\eta_p^2 = 0.060$). Post-hoc comparisons confirmed significantly impaired learning in the $\text{OCD}_{\text{Low-NoSRI}}$

group $OCD_{Low\text{-}NoSRI}$ patients < HCs ($t(183) = 3.87$, $p_{adjusted} = 0.0005$) and $OCD_{Low\text{-}NoSRI}$ patients < $OCD_{HighSRI}$ patients($t(183) = 2.38$, $p_{adjusted} = 0.036$) in session 5 and $OCD_{Low\text{-}NoSRI}$ patients < HCs ($t(183) = 3.41$, $p_{adjusted} = 0.0024$) in session 6 (Figure 3).

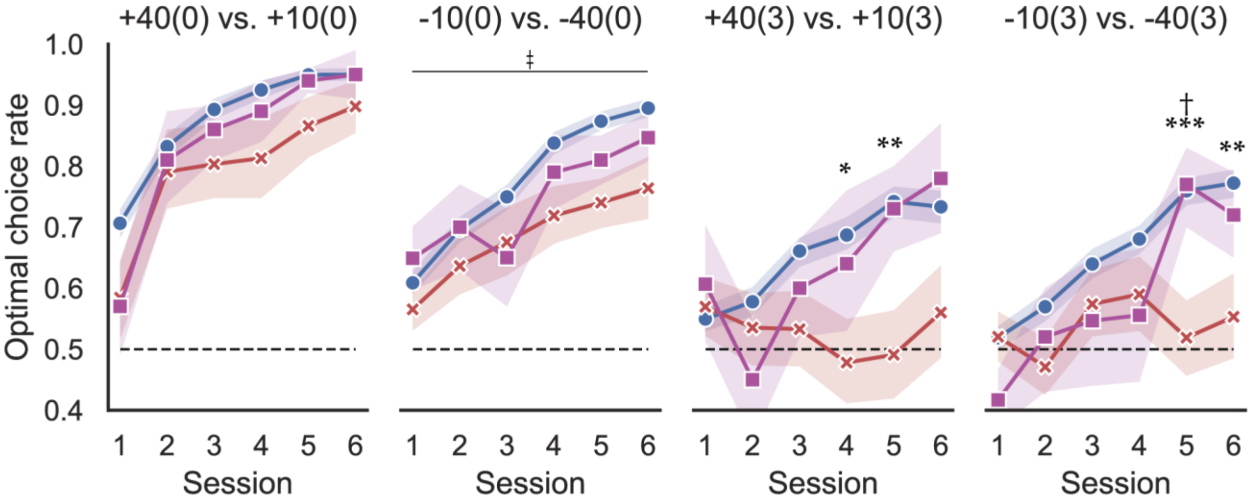

Figure 3. Optimal choice rates of the delayed feedback task. Each panel represents the result for each pair of stimuli [from left to right: pairs with no delays, +40(0) vs. +10(0) and −10(0) vs. −40(0); pairs with delays, +40(3) vs. +10(3) and −10(3) vs. −40(3)]. The line and colored area represented the mean and standard error of the optimal choice rates of groups (red cross, $OCD_{Low\text{-}NoSRI}$ patients; magenta square, $OCD_{HighSRI}$ patients; blue circle, HCs). The horizontal dashed line at 0.5 is the chance level. *, **, and ***$p_{adjusted} < 0.05$, 0.01, and 0.001 with Bonferroni-Holm–corrected post-hoc comparisons of the simple main effects of group ($OCD_{Low\text{-}NoSRI}$ patients < HCs); $^{\dagger}p_{adjusted} < 0.05$, Bonferroni-Holm–corrected post-hoc comparisons of the simple main effects of group ($OCD_{Low\text{-}NoSRI}$ patients < $OCD_{HighSRI}$ patients); $^{\ddagger}p_{adjusted} < 0.05$, Bonferroni-Holm–corrected post-hoc comparisons of the main effect of group ($OCD_{Low\text{-}NoSRI}$ patients < HCs).

Computational model-based behavioral analysis and results

We fitted the behavioral data with the actor-critic learning model using the learning rate (α), exploration-exploitation degree (β), and separation of the eligibility trace for reinforcement and punishment ($\nu^{\pm}$) (see Online Methods). Because we were specifically interested in the individual variance represented by reinforcement learning parameters, we fitted parameters in each participants' data independently using maximum a posteriori estimation rather than pooled data as a group[24].

To clarify the ν^{+}/ν^{-} distribution of each group, we projected each participants' estimated parameter to ν^{+}/ν^{-} space and visualized it using kernel density estimation (Figure 4a). Consistent with our computational model simulation of OCD, $\text{OCD}_{\text{Low-NoSRI}}$ participants were distributed to the $\nu^{+} > \nu^{-}$ imbalanced area, whereas the distribution was balanced in HCs and $\text{OCD}_{\text{HighSRI}}$ participants (Figure 4a). To compare the ν^{+}/ν^{-} distribution among groups, we confirmed the multivariate homogeneity of group dispersions (F(2, 183) = 1.06, p = 0.35) and applied permutational multivariate analysis of variance (PERMANOVA). The ν^{+}/ν^{-} distribution was significantly different (PERMANOVA: F(1, 184) = 6.41, p = 0.0039), with Bonferroni-Holm–corrected post-hoc pairwise comparisons revealing a significant difference between $\text{OCD}_{\text{Low-NoSRI}}$ participants and HCs ($p_{\text{adjusted}} = 0.014$) (Figure 4a). Parameters α and β were not significantly different among groups (Kruskal-Wallis test, $p > 0.05$).

We conducted clustering analysis using hierarchical density-based spatial clustering of applications with noise (HDBSCAN)[25] to evaluate the diversity in HCs. HDBSCAN revealed two clusters in HCs: a balanced ν cluster (Figure 4a, blue circle; n = 83) and an imbalanced ν cluster (Figure 4a, white circle; n = 59); the remaining 11 HCs were not clustered. Obsessive-compulsive trends and a propensity to adhere to fine-grained details were compared between clusters using five subscales of the Padua Inventory (PI; "Checking", "Dirt", "Doubt", "Impulse", "Precision")[26] and the Attention to Detail subscale of the Autism Quotient (AQ)[27], respectively. The imbalanced HC cluster showed significantly higher scores in the PI Checking score (Brunner-Munzel test, statistic = −2.11, p = 0.019) and the AQ Attention to Detail score (Brunner-Munzel test, statistic = −2.79, p = 0.0030) than the balanced HC cluster (Figure 4b). Significant differences were still found after false discovery rate correction for multiple comparisons among scales (PI Checking, $p_{adjusted} = 0.046$; AQ Attention to Detail, $p_{adjusted} = 0.015$). In OCD participants taking SRIs (n = 20), the equivalent dose of SRIs was significantly correlated with the imbalanced settings of ν ($\nu^{+}–\nu^{-}$) [Spearman's rank correlation (20); r = −0.61, p = 0.0045; Figure 4c]. That is, higher doses of SRIs normalized the imbalanced settings of ν in participants with OCD.

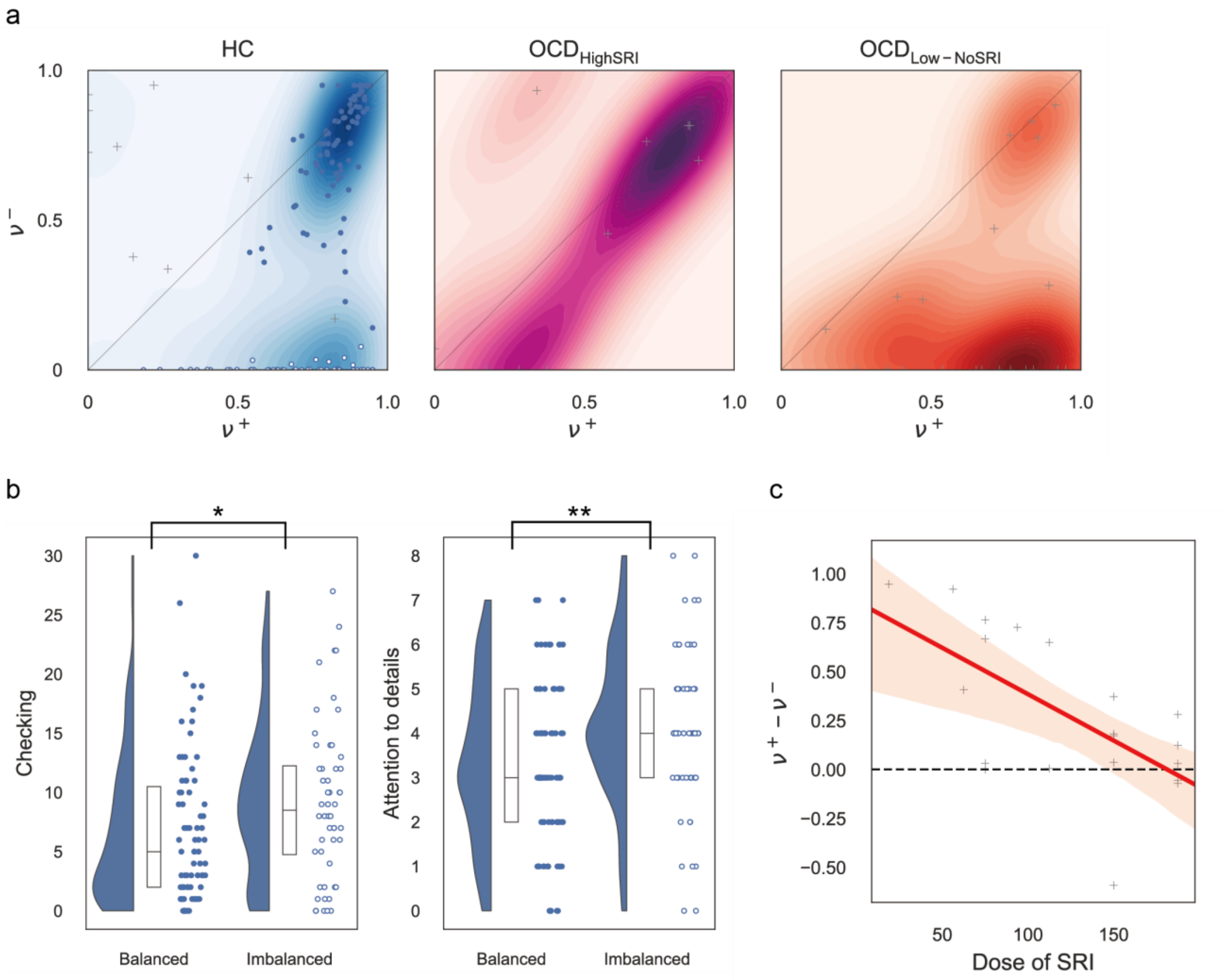

Figure 4. Estimated parameters and their relationships with the clinical characteristics. (a) Distribution of participants in each group projected in the ν^+/ν^- space using the kernel density estimation (blue, HCs; magenta, $\text{OCD}_{\text{HighSRI}}$ patients; red, $\text{OCD}_{\text{Low-NoSRI}}$ patients). The horizontal axis represents ν^+ and the vertical axis represents ν^-. The diagonal line represents the balanced line in the ν^+/ν^- space. Blue and white circles in the leftmost figure represent the balanced and imbalanced clusters in HCs, respectively. Other participants were depicted using +. (b) PI Checking and AQ Attention to Detail scores in the balanced and imbalanced clusters in HCs. The half-violin plot, box plot, and strip plot represent the probability density, interquartile range and median, and raw data, respectively. Blue and white circles represent the balanced and

imbalanced clusters. * and ** represent significant differences (Brunner-Munzel test, * statistic = −2.11, p = 0.019; ** statistic = −2.79, p = 0.0030). (c) Significant correlation between the SRI dose and ν^{+}–ν^{-} [Spearman's rank correlation (20); r = −0.61, p = 0.0045]. The line and colored areas are the regression line and the 95% confidence interval (+: each OCD patient taking SRIs). The dashed line represents the balanced settings of $\nu^{\pm}$.

Neural signatures of OCD and the imbalanced $\nu^{+} > \nu^{-}$ HC cluster

The above behavioral results revealed the presence of imbalanced clusters in both the OCD patients and the HCs. We next explored such imbalanced conditions of $\nu^{+} > \nu^{-}$ in neural circuits. Recently, various behavioral and demographic characteristics and psychiatric conditions have been thought to be represented in a brain network, even in the resting state[14, 15, 28-30]. Here, we explored the neural signature of nonmedicated patients with OCD hypothetically related to $\nu^{+} > \nu^{-}$ and extended it into an imbalanced ($\nu^{+} > \nu^{-}$) HC cluster using rs-fMRI. First, we constructed a whole-brain functional connectivity (FC) matrix using Cole-Anticevic Brain-wide Network Partition (CAB-NP)[31] and compared 49 $\mathrm{OCD}_{\mathrm{NoSRI}}$ patients and 53 HCs (dataset A; Online Methods and Supplementary Table 4) using network-based statistics (NBS)[32] (initial threshold, t = 3.87; 10,000 random permutations; see Online Methods). We detected a significant network component related to OCD (OCD network) with significantly increased connectivity in $\mathrm{OCD}_{\mathrm{NoSRI}}$ patients compared with HCs (p_{adjusted} = 0.022, Figure 5a). The OCD network comprises nodes in the dorsolateral frontal cortex (DLPFC), parietal cortex, retrosplenial cortex, and the hippocampal formation and parahippocampal region. These brain regions mainly belong to the default mode network (DMN) or frontoparietal network (FPN) (Supplementary Figure 1). To confirm the robustness of the OCD network, we compared the mean FC of the OCD network between 10 $\mathrm{OCD}_{\mathrm{NoSRI}}$ patients and 18 HCs in the entirely independent dataset (dataset B; Online Methods and Supplementary Table 5). Similarly, the mean FC within the OCD network was significantly higher in $\mathrm{OCD}_{\mathrm{NoSRI}}$ patients than in HCs (Figure 5b, Brunner-Munzel test, statistic = −3.11, p = 0.0027). Finally, we explored whether the imbalanced ($\nu^{+} > \nu^{-}$) HC cluster showed OCD-like characteristics also regarding a functional network. We compared all FCs of the OCD network between 10 HCs in the imbalanced and balanced clusters (dataset C: independent from dataset A and B; Online Methods and Supplementary Table 6) detected in our delayed feedback task. We found that the FC between the DLPFC and presubiculum was significantly increased in the imbalanced ($\nu^{+} > \nu^{-}$) HC cluster, similar to OCD (Figure 5c, Brunner-Munzel test, statistic = −2.94, p = 0.0051). The FC still showed a significant trend after false discovery rate correction for multiple comparisons among the FCs of the OCD network (p_{adjusted} = 0.066).

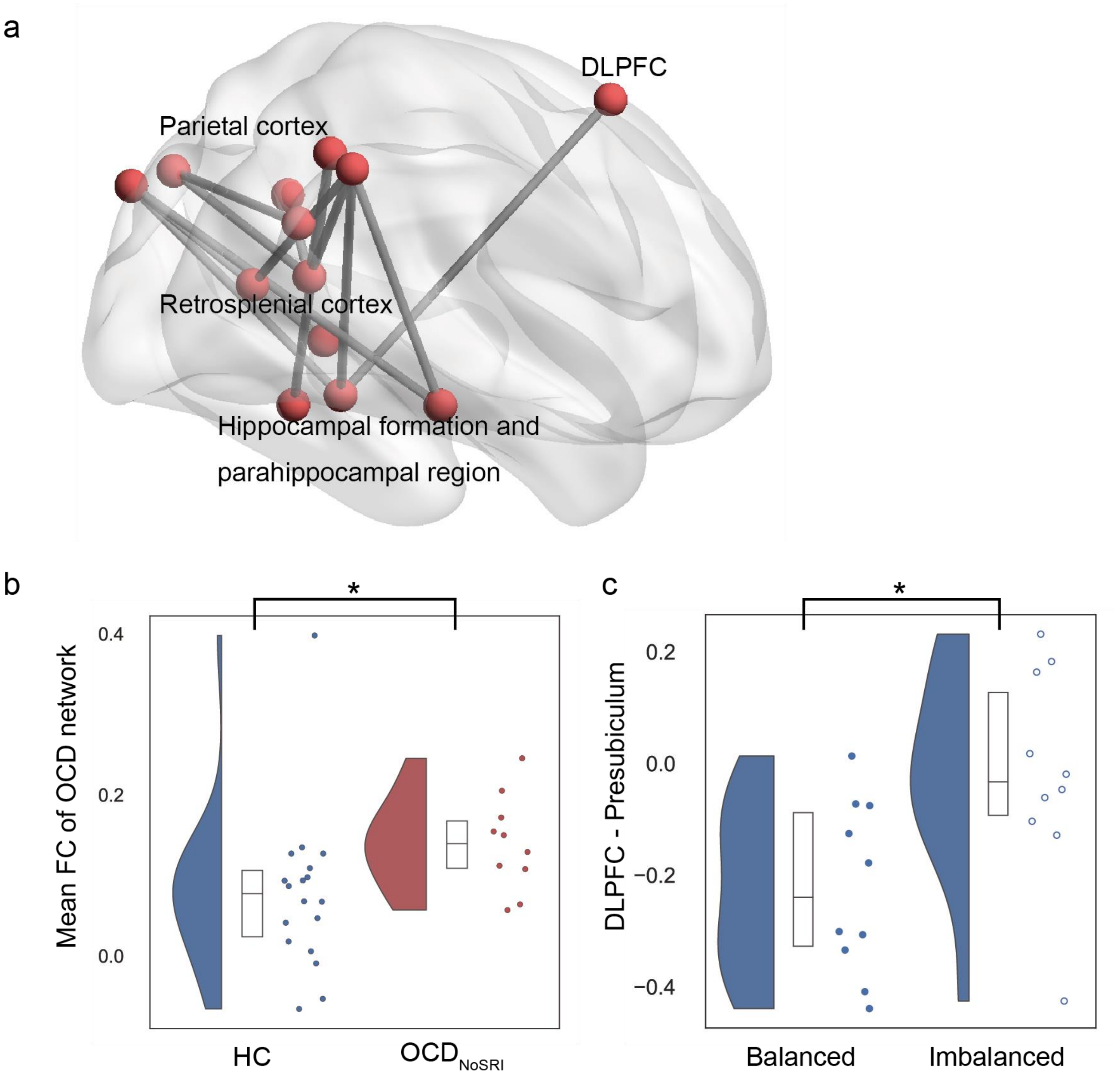

Figure 5. The OCD network and its extension to the imbalanced ($\nu^{+} > \nu^{-}$) HC cluster. There was no overlap of participants among the results in (a), (b), and (c). L and R represent left and right, respectively. (a) Significantly increased functional network component in $\mathrm{OCD}_{\mathrm{NoSRI}}$ participants compared with HCs ($p_{\mathrm{adjusted}} = 0.022$). (b) Significantly increased mean FC within the OCD network of $\mathrm{OCD}_{\mathrm{NoSRI}}$ participants compared with HCs in the completely independent dataset (* Brunner-Munzel test, statistic = −3.11, p = 0.0027). (c) The imbalanced HC cluster showed

OCD-like increased FC between the DLPFC and presubiculum (* Brunner-Munzel test, statistic = −2.94, p = 0.0051). In (b) and (c), the half-violin plot, box plot, and strip plot represent the probability density, interquartile range and median, and raw data, respectively.

Discussion

In this study, we showed that implicit choices in our mind can induce anxiety, which is related to imbalance in eligibility trace time scales for reinforcement and punishment. Specifically, we constructed a computational model of OCD using a separate eligibility trace model (Figure 1a–d) and found extremely imbalanced trace factors $\nu^+ > \nu^-$ in OCD (Figure 4a) and its neural substrate (Figure 5a) in our empirical data. In addition, behavioral therapy (ERP) and psychotropic medication (SRIs), which are the first-line treatments for OCD, were reflected in our computational model (Figure 1d) and behavioral results (Figure 4c), respectively.

While the theoretical framework of the eligibility trace has long been conceptualized in the field of reinforcement learning[4], its empirical evidence has been reported relatively recently[10, 33-35]. The latest theories regarding synaptic plasticity have proposed that the co-activation of pre- and postsynaptic neurons sets a flag at the synapse (eligibility trace) that leads to a weight change only if additional factors (i.e., reinforcement or punishment) are present while the flag is set[12]. These additional factors could be implemented by the phasic activity of some neuromodulators, such as dopamine, which is supposed to represent the prediction error[5]. Although the detailed mechanisms concerning how the trace time scales for reinforcement and punishment are modulated remain unclear, our theoretical consideration of the anxiety-relief transition model showed that the imbalanced trace factors $\nu^+ > \nu^-$ could lead to the condition of repetitive choices of anxiety and its relief in the model, similar to obsession and compulsion in

OCD (Figure 1a–d). Our experimental data from OCD patients and HCs strongly support the predictions from our computational model, namely, the apparent impairment in the learning with delayed feedback (Figure 3) and the extremely imbalanced trace factors $\nu^+ > \nu^-$ in OCD (Figure 4a). It is noteworthy that the imbalanced trace factors $\nu^+ > \nu^-$ are quite convincing because the conventional pathophysiological model of OCD suggests excess tone in the direct pathway over the indirect pathway[13-15], which are supposed to be related to ν^+ and ν^-, respectively.

There have been several computational models of OCD, all of which primarily focused on compulsive behaviors[36-38]. Compulsive actions are thought to be reinforced by the rewarding effect of relief from anxiety[17]. Excessive anxiety will induce excessive reinforcement of compulsion leading to habit formation[38]. This excessively habitual compulsion is considered a cause of OCD[38, 39]. However, obsessive thoughts that drive anxiety also increase with the severity of OCD symptom[17]. Although the essence of OCD symptoms is the abnormal repetition of obsession and compulsion[17], there is so far no unified model to explain the growth of both obsession and compulsion. Our unified model can represent the vicious circle of obsession and compulsion, which are the phenomenological characteristics of OCD[17] (Figure 1a–d). Moreover, our model extends our understanding of the therapeutic effects of first-line treatments of OCD. While ERP seems to promote the appropriate choices to prevent obsession, even under imbalanced trace factors (Figure 1d), SRIs resolve the imbalance itself (Figure 4a and c). In practice guidelines for the treatment of OCD[18], the combination therapy of ERP and SRI is recommended if patients do not respond to ERP monotherapy. We identified such a condition where ERP would fail in our computational model (Figure 1f). SRI add-on therapy in this ERP-refractory condition can be viewed as assisting in the normalization of trace factors. Regarding the relationships between the neuromodulator serotonin and the time scales of the eligibility trace, serotonin seems to modulate the synaptic plasticity in many brain regions directly or indirectly through its regulatory effects on other neuromodulatory systems[12, 33, 40, 41]. In our previous study, we also found similar modulatory effects of the trace factor time scales, that is, depletion of the serotonin precursor tryptophan increased the $\nu^+ > \nu^-$ imbalance compared with the control condition[23]. There is still abundant scope for further research aimed at elucidating how serotonin modulates the time scale of trace factors.

Using rs-fMRI, we found that patients with OCD exhibited significantly increased connectivity of the OCD network component, which consists of the DLPFC, parietal cortex, retrosplenial cortex, hippocampal formation, and parahippocampal region in two independent datasets (Figure 5a and b). Specifically, the OCD network is mainly composed of FCs within the DMN (7 of 13 FCs) and FCs between the DMN and FPN (4 of 13 FCs) (Supplementary Figure 1). These results are quite consistent with a recent meta-analysis of rs-fMRI findings in OCD, in which the authors reported FC alterations within and between the DMN and FPN[42]. Moreover, the DLPFC, parietal cortex, and hippocampus have been implicated in encoding the eligibility traces related to reward-based decision-making for different time scales[12, 43-45]. These brain regions might help to deal with the multiple eligibility trace time scales required for reinforcement learning as a functional network, together with the retrosplenial cortex, which has dense anatomical connections with all of the other detected brain regions[46].

Beyond the range of clinical phenotypes seen in patients with OCD, a broader continuum of the obsessive-compulsive trait is also observed in behavioral and neural characteristics. Specifically, the imbalanced ($\nu^+ > \nu^-$) HC cluster showed a significantly greater propensity for Checking and Attention to Detail (Figure 4b) and increased FC between the DLPFC and presubiculum, which belongs to the OCD network (Figure 5c). These results not only increase the reliability of our clinical findings, but also support the generalizability of our findings to a broader population on the obsessive-compulsive continuum. Further study with greater focus on OCD patients and their unaffected healthy first-degree relatives should be performed to evaluate the potential of our findings as an endophenotype of OCD[47].

In this study, we provided evidence that one's own implicit choices can induce anxiety in OCD. However, anxiety can itself manifest in many different forms. In some situations, anxiety may motivate people to take action or strive to meet a goal. In other situations, people may experience anxiety as a symptom of an anxiety disorder such as social anxiety disorder and generalized anxiety disorder. We previously demonstrated the commonality of different types of anxiety, including obsessive-compulsive anxiety, using fMRI[29]. Because we cannot cover all forms of anxiety in this research, further study with greater focus on the relationships between implicit choices and various representations of anxiety from computational aspects should be performed.

Conclusion

Anxiety can feel like a calamity that befalls us. However, our research shows that our own implicit choices can trigger anxiety. Psychiatric symptoms are often thought of as alterations in the mind that are not directly quantifiable, but they can be directly assessed through the creation of appropriate computational models. Although it is currently difficult to identify treatment-resistant patients from their clinical symptoms, our computational model suggests that patients with extremely imbalanced trace scale parameters may not respond to behavioral therapy alone. These results suggest that our findings could one day be applied to the appropriate selection of OCD treatment. In addition, psychiatric symptoms have been regarded in recent years as a symptom dimension common to various mental diseases, rather than being specific to a disease. In this study, we focused on patients with OCD and healthy participants, but our approach could be applied to the assessment of the anxiety dimension in various populations[48]. Future research is needed to address these hypotheses in prospective longitudinal cohorts or in larger cohorts with various psychiatric symptoms.

Online Methods

Separate eligibility trace model

Eligibility traces determine the weight assignment of outcome prediction errors for recent actions (Figure 1a). If the eligibility traces for positive and negative prediction errors are implemented in distinct neural systems, the respective eligibility traces $\eta_{as}^{+}(t)$ and $\eta_{as}^{-}(t)$ at time step t for action a (a = 0, 1, …) in state s (s = 0, 1, …) obey the following equation[4],

$$\eta_{as}^{\pm}(t+1) = \nu^{\pm}\,\eta_{as}^{\pm}(t) + X_{as}(t),$$

where $X_{as}(t) = 1$ when action a is chosen in state s at time step t and $X_{as}(t) = 0$ otherwise. The factors $\nu^{\pm}$ ($0 < \nu^{\pm} < 1$) determine the decaying time scales of the eligibility traces. The policy parameters $q_{as}(t)$ (a = 0, 1, …) determine choice probability in state s as a soft-max function,

$$P(a|s) = e^{\beta q_{as}(t)} \Big/ \sum_{\acute{a}} e^{\beta q_{\acute{a}s}(t)},$$

where β represents the degree of the exploration-exploitation balance. Each policy parameter is updated as below,

$$q_{as}(t+1) = q_{as}(t) + \alpha[\eta_{as}^{+}(t)\max\{0, \varepsilon(t)\} + \eta_{as}^{-}(t)\min\{0, \varepsilon(t)\}],$$

where $\varepsilon(t)$ denotes the outcome prediction error and $\max\{0, \varepsilon(t)\}$ and $\min\{0, \varepsilon(t)\}$ represent positive and negative components of the prediction error, respectively (Figure 1a). Theoretically, the trace factors $\nu^{\pm}$ should be balanced as $\nu^{+} = \nu^{-}$. However, the separate neural implementation makes it difficult to perfectly maintain the balance. Here, we assumed that the balance could be broken: $\nu^{+} \neq \nu^{-}$.

The outcome prediction error $\varepsilon(t)$ is determined as a function of the current outcome $r(t)$ and the policy parameters $\{q_{as}\}$. The form depends on learning algorithms. In actor-critic learning, the outcome prediction is based on the state value $v_s = \sum_a q_{as}$, and the prediction error $\varepsilon(t) = r(t) + \gamma v_{s(t+1)} - v_{s(t)}$, where $r(t)$ denotes the current outcome and $s(t)$ and $s(t+1)$ denote the current and next states. The parameter γ denotes the discount factor that determines the weight of future prediction. In SARSA and Q-learning, the outcome prediction is based on the action value estimated as each policy parameter q_{as}, and the prediction error $\varepsilon(t) = r(t) + \gamma V_{s(t+1)} - q_{a(t)s(t)}$. The difference is in the term of the next state value: $V_{s(t+1)} = q_{a(t+1)s(t+1)}$ in SARSA and $V_{s(t+1)} = \max_a q_{as(t+1)}$ in Q-learning.

Anxiety-relief transition model

We modeled the mental states involved in anxiety into stochastic transitions between two states: relief (s=0) and anxiety (s=1) (Figure 1b). We assumed two options in each state: "compulsion" (a=1) and "other" (a=0) in the anxiety state, and "obsession" (a=1) and "other" (a=0) in the relief state. We defined a matrix b to determine the state-transition probabilities,

$$P(s(t+1) = 1|a(t) = i, s(t) = j) = b_{ij},$$

$$P(s(t+1) = 0|a(t) = i, s(t) = j) = 1 - b_{ij}.$$

We assumed that every stay in the anxiety state produced a negative outcome normalized to -1. The relative cost of compulsion was denoted as c. Although b_{00}>0 allows passive anxiety, we focused on b_{00}=0 in the main article. More general cases, including passive anxiety, are considered in the Supplementary Note.

Demonstration of OCD-like behavior

For the example in Figure 1, we fixed the parameters of the anxiety-relief transition as b_{00}=0, b_{10}=1, b_{01}=0.9, b_{11}=0.5, and c=0.01 and the learning parameters as α=0.1, β=1, $\gamma = 0.5$, and ν^{+}=0.8, except for ν^{-}. For Figure 1c, the trace factor for punishment was set as ν^{-}=0.6 and the initial values of variables as $\eta_{as}^{\pm}(0) = q_{as}(0) = s(0) = 0$. The simulation consisted of 100,000 time steps. For Figure 1d, we set the trace factors as ν^{-}=0.1 and the initial values as $\eta_{as}^{\pm}(0) = q_{a1}(0) = s(0) = 0$, $q_{00}(0) = 3$, and $q_{10}(0) = -3$, to show that reinforcement of obsession started from even a low obsession rate. After 50,000 time steps, compulsion (a=1 in s=1) was always prevented, and other (a=0 in s=1) was forced regardless of the choice probability, demonstrating the behavioral therapy of ERP.

For Figure 1e and f, we evaluated the fraction of 100 simulation runs in which the obsession rate was reinforced from a low obsession rate on 40 × 40 grids in (ν^{+}, ν^{-}) space. In each simulation, 200 instances of forced obsessions were intermittently caused in the relief state because spontaneous obsession scarcely occurred at a low obsession rate. The simulation was judged to be obsession-reinforced if the choice probability of obsession became larger than the initial value. The initial values of variables were set as $\eta_{as}^{\pm}(0) = q_{a1}(0) = s(0) = 0$, $q_{00}(0) = 5$, and $q_{10}(0) = -5$. Each forced obsession was caused when $\eta_{as}^{\pm}(t)$ sufficiently approached zero in the relief state ($\eta_{as}^{\pm}(t) < 0.001$).

Participants

In total, 33 patients with OCD and 168 HCs participated in the behavioral task. Fifteen HCs were excluded for medical or experimental reasons, and the subsequent analysis was conducted in 33

OCD patients and 153 HCs. There were no significant differences in age, sex, or handedness (Supplementary Table 1). All OCD patients and 13 of the 153 HCs were recruited at Kyoto Prefectural University of Medicine (KPUM), whereas 140 of the 153 HCs were recruited at Advanced Telecommunications Research Institute International (ATR). The Medical Committee on Human Studies at KPUM and the Ethics Committee at ATR approved all procedures in this study. All participants gave written informed consent after receiving a complete description of the study. All methods were carried out following the approved guidelines and regulations. Trained, experienced clinical psychiatrists and psychologists assessed all participants.

All patients were primarily diagnosed using the Structured Clinical Interview for DSM-IV Axis I Disorders-Patient Edition (SCID)[49]. Exclusion criteria were 1) cardiac pacemaker or other metallic implants or artifacts; 2) significant disease, including neurological diseases, disorders of the pulmonary, cardiac, renal, hepatic, or endocrine systems, or metabolic disorders; 3) prior psychosurgery; 4) psychotropic medication except for SRIs; 5) DSM-IV diagnosis of mental retardation and pervasive developmental disorders based on a clinical interview and psychosocial history; and 6) pregnancy. We excluded patients with current DSM-IV Axis I diagnosis of any significant psychiatric illness except OCD as much as possible, and only two patients with trichotillomania, one patient with a tic disorder, one patient with panic disorder, and one patient with bulimia nervosa were included as patients with comorbidities. The experienced clinical psychiatrists or psychologists applied the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS)[21] for clinical evaluation of obsessive-compulsive symptoms in patients with OCD. Handedness was classified based on a modified 25-item version of the Edinburgh Handedness Inventory.

We divided OCD into two groups by medication status. Thirteen patients with OCD were drug-free for all types of psychotropic medication, and the remaining 20 patients were taking SRIs only. The SRI and imipramine equivalent doses[20] are summarized in Supplementary Table 2. OCD patients with higher SRI doses (higher than or equal to imipramine equivalent dose 150 mg) were grouped into the $\text{OCD}_{\text{HighSRI}}$ group (n = 10) and those with lower SRI doses (less than imipramine equivalent dose 150 mg) or no psychotropic medications were grouped into $\text{OCD}_{\text{Low-NoSRI}}$ ($\text{OCD}_{\text{LowSRI}}$, n = 10; $\text{OCD}_{\text{NoSRI}}$, n = 13). The threshold was determined by considering a meta-analysis of the SRI treatment[19] and common clinical doses used in Japan. To evaluate the antidepressant effects of SRI, we compared the depressive symptoms evaluated by the HDRS[22] between $\text{OCD}_{\text{HighSRI}}$ and $\text{OCD}_{\text{Low-NoSRI}}$ patients. There was one missing value of the HDRS in the $\text{OCD}_{\text{HighSRI}}$ group.

Behavioral task and statistical analysis

The delayed feedback task was similar to the one in our previous research[23], except for the presented stimuli. Participants chose one of the two options (abstract cues) displayed on the screen by pressing a left or right button in each trial within 1 s after an auditory cue (Figure 2a). Depending on the selected stimulus, monetary feedback with different outcomes (+10, +40, −10, or −40 yen) was displayed either immediately after the button press or three trials later (Figure 2b and c). We did not offer monetary feedback for the first five trials because participants can learn the relationships between stimuli and outcomes or delays quickly. If participants pressed a button before the auditory cue or more than 1 s passed without any button press, −50 yen was displayed as a punishment. Such trials were considered error trials, and delayed outcomes were not considered.

At each trial, two abstract cues were displayed side by side on the screen (Figure 2a). We prepared 16 pairs (Figure 2b), counterbalancing the number of appearances of each stimulus. The 16 pairs of stimuli were presented in pseudorandom order. Each pair was presented as the scheduled number of trials in each session: each of six pairs [+10(0) vs. +40(0); +10(3) vs. +40(3); −10(0) vs. −40(0); −10(3) vs. −40(3); +10(0) vs. +40(3); −10(0) vs. −40(3)] was presented in 10 trials during a single session, and each of 10 pairs [+40(0) vs. +40(3); +10(0) vs. +10(3); −10(0) vs. −10(3); −40(0) vs. −40(3); +10(3) vs. +40(0); −10(3) vs. −40(0); +10(0) vs. −10(0); +40(0) vs. −40(0); +10(3) vs. −10(3); +40(3) vs. −40(3)] was presented in five trials during a single session. Each participant performed 110 trials during a single session and six sessions in each experiment. About 28 min was required for participants to complete six sessions. At the beginning of each session, the session number was displayed on the screen for 2.5 s. Before the task, each participant practiced the test session under the same task settings except for stimuli, and we confirmed that all participants understood the task set.

The total outcome (except for punishments related to button press errors), reaction time, and the number of error trials were compared between the OCD and HC groups. Patients with OCD showed a significantly lower total monetary outcome compared with HCs [median (interquartile range): OCD patients, 2890 (1590–3600) yen; HCs, 4550 (3020–5950) yen; Brunner-Munzel test, statistic = 4.08, $p = 0.0002$], whereas there were no group differences in reaction time [median (interquartile range): OCD patients, 524.6 (487.1–565.4) ms; HCs, 522.6 (475.3–575.9) ms; Brunner-Munzel test, $p > 0.05$)] and number of button press errors [median (interquartile range): OCD patients, 4 (2–9); HC, 4 (2–8); Brunner-Munzel test, $p > 0.05$). These results suggested impaired learning in OCD individuals compared with HCs. Based on our hypothesis of trace factors ν^+/ν^-, patients with OCD would show impaired learning in the stimuli with delayed feedback. Therefore, the optimal choice rate of the representative four pairs of stimuli with the same delay and different magnitude [pairs with no delays: +40(0) vs. +10(0) and −10(0) vs. −40(0); pairs with delays: +40(3) vs. +10(3) and −10(3) vs. −40(3)] were compared among groups. A mixed-design two-way repeated-measures ANOVA with a within-participants factor of sessions (1–6 sessions) and a between-participants factor of groups ($\mathrm{OCD}_{\mathrm{Low\text{-}NoSRI}}$ patients, $\mathrm{OCD}_{\mathrm{HighSRI}}$ patients, and HCs) was conducted to clarify the between-group differences. Degrees of freedom were corrected using Chi-Muller's epsilon because Mendoza's multisample sphericity test indicated that the assumption of sphericity had been violated. Bonferroni-Holm–corrected post-hoc comparisons were conducted to clarify the between-group differences.

Model comparison and parameter estimation for the delayed feedback task

We fitted the behavioral data with actor-critic learning with separate eligibility traces for positive and negative prediction errors. We defined 16 states for possible pairs of stimuli presented at each trial. Available actions at each state involved the choosing of alternative stimuli. To facilitate model-fitting in the face of limited experimental data from each participant, we used regularizing priors that favored realistic values and maximum a posteriori estimation rather than maximum likelihood estimation[24]. The learning rate α and trace factor ν were constrained to the range of $0 \leq \alpha \leq 0.95$ and $0 \leq \nu \leq 0.95$ with a uniform prior. The exploration-exploitation degree β was constrained to the range of $0 \leq \beta \leq 100$ with a gamma (2,3) prior distribution that favored relatively lower values. We fixed the discount factor $\gamma = 0$, because the term of the next state value was just noise in our delayed feedback task in which the state of each trial was randomly selected. We optimized parameters by minimizing the negative log posterior of the data with different parameter settings using the hyperopt package[50]. Likelihood ratio tests to assess the contribution of the additional parameter in our model (four parameters: α, β, ν^+, ν^-) compared with the standard actor-critic learning model (three parameters: α, β, ν) showed that the additional parameter was justified in 128 of 186 participants (X^2 test with one degree of freedom, $p < 0.05$). Because our target model was validated in behavioral data, we applied the separate eligibility trace model in the substantive analysis.

We evaluated the performance of parameter estimation in actor-critic learning with separate eligibility traces. We created 200 simulation data points using a model with the following parameters: α, 0.1 ± 0.05 (mean ± SD); β, 1 ± 0.2; ν^+ and ν^-, randomly selected within 0.01–0.95. The parameter estimation was quite accurate[51] (Supplementary Figure 2). Specifically, Pearson's r and the mean absolute error between the true and estimated ν^+ or ν^- were 0.99 and 0.03, respectively (Supplementary Figure 2).

We compared the ν^+/ν^- distribution in our model among three groups ($\text{OCD}_{\text{Low-NoSRI}}$ patients, $\text{OCD}_{\text{HighSRI}}$ patients, and HCs) using PERMANOVA with the ADNOIS function and 10,000 permutations using the Euclidean distance implemented in the statistical package R[52]. The multivariate homogeneity of group dispersions was confirmed with the *betadisper* function with 10,000 permutations in R[52]. The learning rate α and inverse temperature β were compared using the Kruskal-Wallis test. To further investigate relationships between clinical characteristics and estimated parameters in the HC group, we conducted clustering analysis using HDBSCAN[25]. We detected two clusters (balanced cluster, n = 83; imbalanced cluster, n = 59; the remaining 11 HCs were not clustered) and evaluated their obsessive-compulsive trait using the five PI subscales: "Checking", "Dirt", "Doubt", "Impulse", "Precision"[26]. In addition, the propensity to adhere to fine-grained details was evaluated using the Attention to Detail subscale of AQ[27]. There were 15 missing values in PI (n = 127) and 1 missing value in AQ (n = 141). The therapeutic effects of SRIs were evaluated by the Spearman's rank correlation between the SRI dose and the imbalanced settings of ν ($\nu^+ - \nu^-$).

Imaging data acquisition, preprocessing, and statistical analysis

rs-fMRI data were collected using three different MRI scanners: 49 $\mathrm{OCD}_{\mathrm{NoSRI}}$ participants and 53 HCs at Kajiicho Medical Imaging Center (dataset A), 10 $\mathrm{OCD}_{\mathrm{NoSRI}}$ participants and 18 HCs at Kyoto Prefectural University of Medicine (dataset B) for replication of the findings of dataset A, and 20 HCs with the delayed feedback task (10 HCs for the imbalanced cluster and the remaining 10 HCs for the balanced cluster) in ATR (dataset C). There was no overlap of participants among datasets. All demographic distributions were matched between groups (dataset A, Supplementary Table 4; dataset B, Supplementary Table 5; dataset C, Supplementary Table 6). Some of the participants in dataset A and B were included in our previous studies conducted for a different purpose using a different method[14, 15, 28, 29]. All fMRI imaging protocols using gradient EPI sequences are summarized in Supplementary Table 3. High-resolution T1-weighted structural images were also acquired.

Preprocessing of rs-fMRI data was conducted using fmriprep_ciftify 1.3.0.post2-2.3.0[53, 54]. Briefly, typical preprocessing steps such as slice timing correction, motion correction, and spatial normalization into Montreal Neurological Institute space were conducted using fmriprep. We then converted the data from the volumetric NIfTI format to the surface-based CIFTI format (https://www.nitrc.org/projects/cifti/) using ciftify[54]. To remove artifacts and increase the signal/noise ratio, the time course of the rs-fMRI data was detrended, bandpass-filtered (0.01–0.08 Hz), and linearly regressed out of nuisance variables (the temporal fluctuations of the entire brain and six head motion parameters, their derivatives, and six principal components of anatomical CompCor[55]). With respect to motion artifacts, framewise displacement (FD) was not significantly different between groups in all datasets [median (interquartile range); dataset A: OCD patients, 0.082 (0.070–0.097) mm; HCs, 0.087 (0.067–0.10) mm; Brunner-Munzel test, $p > 0.05$; dataset B: OCD patients, 0.095 (0.067–0.12) mm; HCs, 0.079 (0.062–0.094) mm; Brunner-Munzel test, $p > 0.05$; dataset C: balanced cluster, 0.13 (0.11–0.18) mm; imbalanced cluster, 0.11 (0.094–0.14) mm; Brunner-Munzel test, $p > 0.05$]. The first six functional scans were discarded to allow magnetization to reach equilibrium. For each participant, mean time series were extracted from 360 cortical and 358 subcortical parcels using CAB-NP[31], which is the comprehensive whole-brain solution for large-scale functional networks based on the cortical parcellation developed by Glasser et al.[56] (Human Connectome Project Multi-Modal Parcellation). Pearson correlation coefficients were calculated between each pair of parcels and transformed to Fisher's Z scores to obtain the FC matrix.

To evaluate the neural substrate of nonmedicated patients with OCD hypothetically related to $\nu^+ > \nu^-$, we compared the FC matrices between 49 $\mathrm{OCD}_{\mathrm{NoSRI}}$ patients and 53 HCs (dataset A) using NBS[32]. We chose NBS because it facilitates the detection of a subnetwork related to the condition of interest while controlling the family-wise error rate. Briefly, NBS was performed in the following two steps. First, the between-group comparison for every possible FC was conducted and thresholded at $t = 3.87$ (corresponding to $p = 0.0001$) by considering the FD as a nuisance variable. We detected the thresholded subnetworks (networks of nodes interconnected by significant FCs), and their network size was calculated. Second, the significance of subnetworks was tested using 10,000 random permutations of groups, which determined the null distribution of the largest subnetwork size. Only subnetworks whose network size exceeded the estimated family-wise error-corrected p-value 0.05 were identified as the network that was significantly different between OCD$_{\mathrm{NoSRI}}$ patients and HCs. The detected subnetwork was visualized using BrainNet Viewer[57]. To confirm the robustness of the OCD network, we compared the mean FC of the detected OCD network between 10 OCD$_{\mathrm{NoSRI}}$ patients and 18 HCs in the entirely independent dataset (dataset B) using the Brunner-Munzel test. To further explore whether the imbalanced ($\nu^{+} > \nu^{-}$) HC cluster showed OCD-like characteristics also regarding a functional network, we compared every FC of the OCD network between 10 HCs each in the imbalanced and balanced clusters (dataset C) detected in our delayed feedback task using the Brunner-Munzel test.

Data Availability

The patients' data supporting the conclusion of this paper are not publicly available due to them containing information that could compromise research participant privacy or consent. The theoretical derivation of our computational model is in the Supplementary Note along with the MATLAB source code.

Acknowledgments

All authors had full access to the data used in this study and the corresponding author S.C. Tanaka takes responsibility for the integrity of the data and the accuracy of the data analysis. These research results have been achieved by "Research and development of technology for enhancing functional recovery of elderly and disabled people based on non-invasive brain imaging and robotic assistive devices", the Commissioned Research of National Institute of Information and Communications Technology (NICT), JAPAN (S.C. Tanaka). The followings partly supported this study: JSPS KAKENHI Grant Number JP16K01958 and JP16H06396 (S.C. Tanaka), JP16H01516 and JP18H05524 (Yutaka Sakai); 16H01512 and the Sakamoto Research Foundation of Psychiatric Diseases (Yuki Sakai); the Joint Usage/Research Center ("Behavioral economics") of Institute of Social and Economic Research, Osaka University (J. Narumoto).

Author Contributions

Yuki S., Yutaka S., J.N., and S.C.T designed the study. Yutaka S. developed the theory and performed computational modeling. Yuki S., Y.A., J.N., and S.C.T. collected the data. Yuki S. conducted the computational modeling and statistical analysis of the data. Yuki S. and Yutaka S.

wrote the manuscript, which was edited by all authors. S.C.T. acquired funding to support theory development and data analysis.

Competing Interests Statement

The authors declare no competing financial interests.

References

1. Karni, A. & Sagi, D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc Natl Acad Sci U S A* **88**, 4966-4970 (1991).

2. Seitz, A.R. & Watanabe, T. Psychophysics: Is subliminal learning really passive? *Nature* **422**, 36 (2003).

3. Wichers, M., *et al.* From affective experience to motivated action: tracking reward-seeking and punishment-avoidant behaviour in real-life. *PLoS One* **10**, e0129722 (2015).

4. Sutton, R.S. & Barto, A.G. Reinforcement Learning: An Introduction. (MIT Press, Cambridge, MA. 1998).

5. Schultz, W., Dayan, P. & Montague, P.R. A neural substrate of prediction and reward. *Science* **275**, 1593-1599 (1997).

6. Graybiel, A.M. The basal ganglia. *Curr Biol* **10**, R509-511 (2000).

7. Kravitz, A.V., Tye, L.D. & Kreitzer, A.C. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat Neurosci* **15**, 816-818 (2012).

8. Nonomura, S., *et al.* Monitoring and updating of action selection for goal-directed behavior through the striatal direct and indirect pathways. *Neuron* **99**, 1302-1314. e1305 (2018).

9. Shen, W., Flajolet, M., Greengard, P. & Surmeier, D.J. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* **321**, 848-851 (2008).

10. Yagishita, S., *et al.* A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* **345**, 1616-1620 (2014).

11. Izhikevich, E.M. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb Cortex* **17**, 2443-2452 (2007).

12. Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D. & Brea, J. Eligibility traces and plasticity on behavioral time scales: experimental support of neoHebbian three-factor learning rules. *Frontiers in neural circuits* **12**, 53 (2018).

13. Pauls, D.L., Abramovitch, A., Rauch, S.L. & Geller, D.A. Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. *Nat Rev Neurosci* **15**, 410-424 (2014).

14. Sakai, Y., *et al.* Corticostriatal functional connectivity in non-medicated patients with obsessive-compulsive disorder. *Eur Psychiatry* **26**, 463-469 (2011).

15. Abe, Y., *et al.* Hyper-influence of the orbitofrontal cortex over the ventral striatum in obsessive-compulsive disorder. *Eur Neuropsychopharmacol* **25**, 1898-1905 (2015).

16. Salkovskis, P.M. Understanding and treating obsessive-compulsive disorder. *Behav Res Ther* **37 Suppl 1**, S29-52 (1999).

17. Cavedini, P., Gorini, A. & Bellodi, L. Understanding obsessive-compulsive disorder: focus on decision making. *Neuropsychol Rev* **16**, 3-15 (2006).

18. Koran, L.M., *et al.* Practice guideline for the treatment of patients with obsessive-compulsive disorder. *Am J Psychiatry* **164**, 5-53 (2007).

19. 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).

20. Inada, T. & Inagaki, A. Psychotropic dose equivalence in Japan. *Psychiatry Clin Neurosci* **69**, 440-447 (2015).

21. Nakajima, T., *et al.* Reliability and validity of the Japanese version of the Yale-Brown Obsessive-Compulsive Scale. *Psychiatry Clin Neurosci* **49**, 121-126 (1995).

22. Hamilton, M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* **6**, 278-296 (1967).

23. Tanaka, S.C., *et al.* Serotonin affects association of aversive outcomes to past actions. *J Neurosci* **29**, 15669-15674 (2009).

24. Daw, N.D. Trial-by-trial data analysis using computational models. in *Decision Making, Affect, and Learning: Attention and performance XXIII* (eds. Delgado, M.R., Phelps, E.A. & Robbins, T.W.) 3-38 (Oxford Scholarship Online, Oxford, 2011).

25. Campello, R.J., Moulavi, D., Zimek, A. & Sander, J. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM T Knowl Discov D* **10**, 5 (2015).

26. Sugiura, Y. & Tanno, Y. Self-report inventory of obsessive-compulsive symptoms: Reliability and validity of the Japanese version of the Padua Inventory. *Arch Pschyiatr Diagn Clin Eval* **11**, 175-189 (2000).

27. Wakabayashi, A., Tojo, Y., Baron-Cohen, S. & Wheelwright, S. [The Autism-Spectrum Quotient (AQ) Japanese version: evidence from high-functioning clinical group and normal adults]. *Shinrigaku Kenkyu* **75**, 78-84 (2004).

28. Takagi, Y., *et al.* A neural marker of obsessive-compulsive disorder from whole-brain functional connectivity. *Sci Rep* **7**, 7538 (2017).

29. Takagi, Y., *et al.* A common brain network among state, trait, and pathological anxiety from whole-brain functional connectivity. *Neuroimage* **172**, 506-516 (2018).

30. Smith, S.M., *et al.* A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat Neurosci* **18**, 1565-1567 (2015).

31. Ji, J.L., *et al.* Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage* **185**, 35-57 (2019).

32. Zalesky, A., Fornito, A. & Bullmore, E.T. Network-based statistic: identifying differences in brain networks. *Neuroimage* **53**, 1197-1207 (2010).

33. He, K., *et al.* Distinct eligibility traces for LTP and LTD in cortical synapses. *Neuron* **88**, 528-538 (2015).

34. Brzosko, Z., Schultz, W. & Paulsen, O. Retroactive modulation of spike timing-dependent plasticity by dopamine. *Elife* **4**, e09685 (2015).

35. Brzosko, Z., Zannone, S., Schultz, W., Clopath, C. & Paulsen, O. Sequential neuromodulation of Hebbian plasticity offers mechanism for effective reward-based navigation. *Elife* **6**, e27756 (2017).

36. Hauser, T.U., *et al.* Increased decision thresholds enhance information gathering performance in juvenile obsessive-compulsive disorder (OCD). *PLoS Comput Biol* **13**, e1005440 (2017).

37. Vaghi, M.M., *et al.* Compulsivity reveals a novel dissociation between action and confidence. *Neuron* **96**, 348-354.e4 (2017).

38. Gillan, C.M. & Robbins, T.W. Goal-directed learning and obsessive-compulsive disorder. *Philos Trans R Soc Lond B Biol Sci* **369** (2014).

39. Gillan, C.M. & Sahakian, B.J. Which is the driver, the obsessions or the compulsions, in OCD? *Neuropsychopharmacology* **40**, 247-248 (2015).

40. Brzosko, Z., Mierau, S.B. & Paulsen, O. Neuromodulation of spike-timing-dependent plasticity: past, present, and future. *Neuron* **103**, 563-581 (2019).

41. Cavaccini, A., *et al.* Serotonergic signaling controls input-specific synaptic plasticity at striatal circuits. *Neuron* **98**, 801-816 e807 (2018).

42. Gursel, D.A., Avram, M., Sorg, C., Brandl, F. & Koch, K. Frontoparietal areas link impairments of large-scale intrinsic brain networks with aberrant fronto-striatal interactions in OCD: a meta-analysis of resting-state functional connectivity. *Neurosci Biobehav Rev* **87**, 151-160 (2018).

43. Ambrose, R.E., Pfeiffer, B.E. & Foster, D.J. Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron* **91**, 1124-1136 (2016).

44. Gersch, T.M., Foley, N.C., Eisenberg, I. & Gottlieb, J. Neural correlates of temporal credit assignment in the parietal lobe. *PLoS One* **9**, e88725 (2014).

45. Asaad, W.F., Lauro, P.M., Perge, J.A. & Eskandar, E.N. Prefrontal neurons encode a solution to the credit-assignment problem. *J Neurosci* **37**, 6995-7007 (2017).

46. Vann, S.D., Aggleton, J.P. & Maguire, E.A. What does the retrosplenial cortex do? *Nat Rev Neurosci* **10**, 792-802 (2009).

47. Chamberlain, S.R. & Menzies, L. Endophenotypes of obsessive-compulsive disorder: rationale, evidence and future potential. *Expert Rev Neurother* **9**, 1133-1146 (2009).

48. Gillan, C.M., *et al.* Comparison of the association between goal-directed planning and self-reported compulsivity vs obsessive-compulsive disorder diagnosis. *JAMA Psychiatry*, 1-10 (2019).

References for Online Methods and Supplementary Information

49. First, M.B., Spizer, R.L., Gibbon, M. & Williams, J.B.W. *Structures Clinical Interview for Axis I DSM-IV Disorders - Patient Edition (CID-I/P)* (Biometrics Research Department, New York State Psychiatric Institute, New York, 1994).

50. Bergstra, J.S., Bardenet, R., Bengio, Y. & Kégl, B. Algorithms for hyper-parameter optimization. in *Advances in Neural Information Processing Systems* (eds. Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F. & Weinerger, K.Q.) 2546-2554 (Neural Information Processing Systems, Inc., San Diego, 2011).

51. Yazdani, S., Vahabie, A.-H., Araabi, B.N. & Ahmadabadi, M.N. Better than maximum likelihood estimation of model-based and model-free learning style. Preprint at https://www.biorxiv.org/content/10.1101/296335v1 (2018).

52. Anderson, M.J., Walsh, D.C.I., Robert Clarke, K., Gorley, R.N. & Guerra-Castro, E. Some solutions to the multivariate Behrens-Fisher problem for dissimilarity-based analyses. *Aust NZ J Stat* **59**, 57-79 (2017).

53. Esteban, O., *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111-116 (2019).

54. Dickie, E.W., *et al.* Ciftify: A framework for surface-based analysis of legacy MR acquisitions. *Neuroimage* **197**, 818-826 (2019).

55. Behzadi, Y., Restom, K., Liau, J. & Liu, T.T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90-101 (2007).

56. Glasser, M.F., *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171-178 (2016).

57. Xia, M., Wang, J. & He, Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One* **8**, e68910 (2013).