PLOS BIOLOGY

RESEARCH ARTICLE

Individuals with problem gambling and obsessive-compulsive disorder learn through distinct reinforcement mechanisms

Shinsuke Suzuki[1,2]*, **Xiaoliu Zhang**[3], **Amir Dezfouli**[4], **Leah Braganza**[3], **Ben D. Fulcher**[5], **Linden Parkes**[3,6], **Leonardo F. Fontenelle**[3], **Ben J. Harrison**[7], **Carsten Murawski**[1], **Murat Yücel**[3‡], **Chao Suo**[3‡]

1 Centre for Brain, Mind and Markets, The University of Melbourne, Carlton, Australia, **2** Center for the Promotion of Social Data Science Education and Research, Hitotsubashi University, Tokyo, Japan, **3** BrainPark, Turner Institute for Brain and Mental Health, School of Psychological Sciences, and Monash Biomedical Imaging Facility, Monash University, Clayton, Australia, **4** Data61, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Sydney, Australia, **5** School of Physics, The University of Sydney, Sydney, Australia, **6** Department of Bioengineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **7** Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne, Carlton, Australia

‡ These authors are joint senior authors on this work.
* shinsuke.szk@gmail.com

Check for updates

OPEN ACCESS

Citation: Suzuki S, Zhang X, Dezfouli A, Braganza L, Fulcher BD, Parkes L, et al. (2023) Individuals with problem gambling and obsessive-compulsive disorder learn through distinct reinforcement mechanisms. PLoS Biol 21(3): e3002031. https://doi.org/10.1371/journal.pbio.3002031

Academic Editor: Thorsten Kahnt, National Institute on Drug Abuse Intramural Research Program, UNITED STATES

Received: July 11, 2022

Accepted: February 8, 2023

Published: March 14, 2023

Copyright: © 2023 Suzuki et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data is available from BrainPark at Monash University (BrainPark@monash.edu), conditioned upon the approval of the Human Research Ethics Committee of Monash University. The summary data and code necessary to reproduce the analyses shown in the paper are available at https://osf.io/v7em5/.

Funding: This work was supported by the National Health and Medical Research Council of Australia (APP236175 to M.Y.). The funder had no role in

Abstract

Obsessive-compulsive disorder (OCD) and pathological gambling (PG) are accompanied by deficits in behavioural flexibility. In reinforcement learning, this inflexibility can reflect asymmetric learning from outcomes above and below expectations. In alternative frameworks, it reflects perseveration independent of learning. Here, we examine evidence for asymmetric reward-learning in OCD and PG by leveraging model-based functional magnetic resonance imaging (fMRI). Compared with healthy controls (HC), OCD patients exhibited a lower learning rate for worse-than-expected outcomes, which was associated with the attenuated encoding of negative reward prediction errors in the dorsomedial prefrontal cortex and the dorsal striatum. PG patients showed higher and lower learning rates for better- and worse-than-expected outcomes, respectively, accompanied by higher encoding of positive reward prediction errors in the anterior insula than HC. Perseveration did not differ considerably between the patient groups and HC. These findings elucidate the neural computations of reward-learning that are altered in OCD and PG, providing a potential account of behavioural inflexibility in those mental disorders.

Introduction

A central challenge in computational psychiatry is elucidating the fundamental brain processes underlying mental disorders [1–4]. By linking formal models of behaviour with neural data, researchers have started to uncover the basic neural mechanisms of psychiatric symptoms such as anhedonia [5,6] and hallucination [7].

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: M.Y. receives funding from National Health and Medical Research Council of Australia schemes, Monash University, and Australian Government funding bodies such as the Australian Research Council, Australian Defence Science and Technology, and the Department of Industry, Innovation and Science. He has also received philanthropic donations from the David Winston Turner Endowment Fund, Wilson Foundation, as well as payments in relation to court-, expert witness-, and/or expert review-reports. Finally, he has received funding to conduct sponsored Investigator-Initiated trials (including Incannex Healthcare Ltd). He also sits on the Advisory Boards of Centre of The Urban Mental Health, University of Amsterdam; Enosis Therapeutics; and Monash Biomedical Imaging Centre. L.P. was supported by the National Institute Of Mental Health of the National Institutes of Health (K99MH127296) and a 2020 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. The other authors declare no competing interests.

Abbreviations: ACC, anterior cingulate cortex; AIC, Akaike information criterion; BF, Bayes factor; fMRI, functional magnetic resonance imaging; GLMM, generalised linear mixed model; HC, healthy control; HDI, highest density interval; mPFC, medial prefrontal cortex; OCD, obsessive-compulsive disorder; OCI-R, obsessive-compulsive inventory-revised; PG, pathological gambling; RL, reinforcement learning; ROI, regions of interest; SSRI, selective serotonin reuptake inhibitor; WAIC, widely applicable Akaike information criterion.

Learning the value of a course of action is critical for appropriate decision-making [8]. The reinforcement learning (RL) account of decision-making posits that such learning is driven by reward prediction errors [9]. The errors are defined as the discrepancy between the actual and expected reward. They are encoded in a network of brain regions, including the midbrain [10,11], striatum [12,13], medial prefrontal cortex (mPFC) [14,15], and anterior insula [16,17] (see [18] for a recent meta-analysis). Apart from clarifying our basic understanding of decision-making, recent studies in computational psychiatry have explored the brain mechanisms related to psychiatric symptoms based on RL frameworks [19–23].

Both obsessive-compulsive disorder (OCD) and pathological gambling (PG) are characterised by deficits in behavioural flexibility [24–30]. For example, a core symptom of OCD is repetitive behaviours (e.g., excessive hand washing and endless rechecking). Individuals with PG experience difficulty in stopping gambling behaviour despite known adverse consequences. These symptoms may be interpreted as a manifestation of the concept of behavioural addiction [31]. Furthermore, our previous study identified common dimensions (i.e., disinhibition, impulsivity, and compulsivity) of various symptoms across the two disorders, which are located on a continuum with healthy participants [32]. On the other hand, the concept of behavioural addiction remains controversial despite the similarities between PG and OCD. Indeed, DSM-5 classifies gambling disorder as an addictive disorder that is separate from OCD and impulse-control disorders [27,30,33,34]. Whether common or distinct neurocomputational processes mediate the behavioural inflexibility in OCD and PG remains elusive.

One influential explanation suggests that behavioural inflexibility can reflect perseveration independent of the outcomes of decisions [35,36]. That is, inflexibility results from a tendency to repeat the same action, regardless of its outcome. In other words, the same action is repeated even though it caused harmful consequences in the past. Yet, previous studies in computational psychiatry have often failed to detect increased perseveration in OCD and PG patients [37–41]. An alternative account posits that inflexibility reflects asymmetric RL, i.e., overlearning from better-than-expected outcomes and underlearning from worse-than-expected outcomes [42,43]. In formal terms, the learning rate—the extent to which new information (e.g., reward, loss, and neutral outcome) modulates future behaviour—is different for value updating from positive and negative reward prediction errors. This asymmetric learning results in excessive reinforcement and deficient devaluation of the current action, leading to inflexibility of choice [43]. However, examinations of the asymmetric RL account in OCD and PG remain sparse.

In this study, we hypothesised that neural computations underlying RL are altered in OCD and PG. There is supporting evidence for this hypothesis. Both OCD and PG are associated with altered processing of reward and loss information in a number of brain regions, including the striatum, mPFC, adjacent anterior cingulate cortex (ACC), and insula [44–48] (but see [49]). Two recent studies have combined a formal RL model with a neuroimaging experiment to demonstrate an abnormal signal of reward prediction error in the ACC [37,50]. Furthermore, the neurotransmitter dopamine, which signals reward prediction errors in the context of RL, has been implicated in the physiopathology of OCD and PG [50–52].

To test the hypothesis, we recruited 29 patients with OCD, 17 patients with PG, and 34 healthy controls (HCs). The three groups of participants performed a probabilistic instrumental learning task while undergoing MRI (Fig 1A). The experimental task consisted of two sessions, with three conditions (reward, avoidance, and neutral trials) interleaved across trials in a pseudo-randomised order. That is, one of the conditions was presented at random on each trial. Participants repeatedly made choices to earn rewards in the reward trials and to avoid losses in the avoidance trials. On each trial, they selected one of two stimuli, and then received a reward/loss or nothing depending on the probability assigned to the chosen option (Fig 1A).

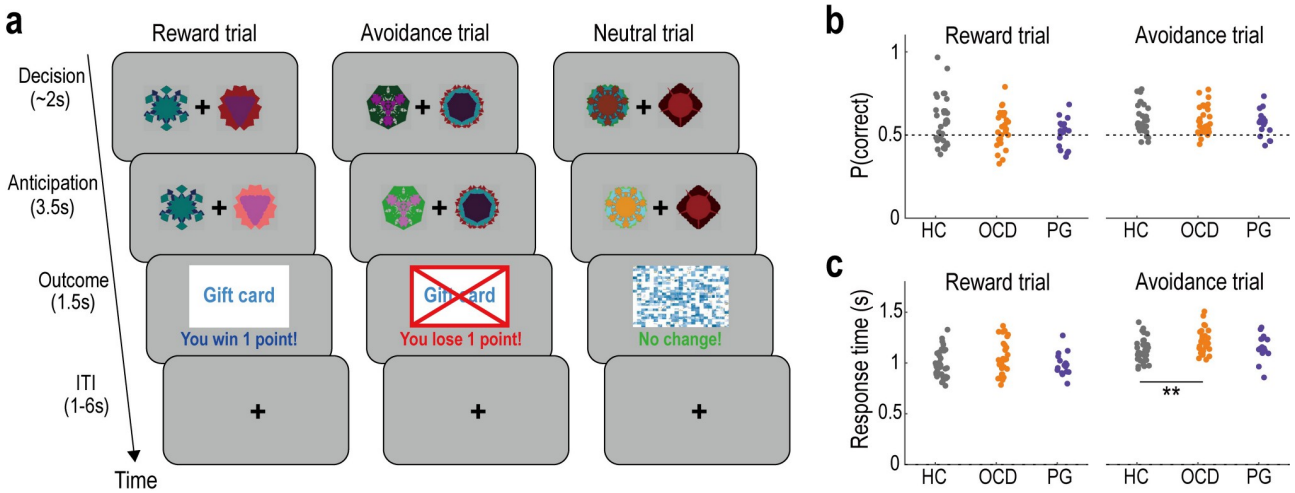

Fig 1. Experimental task and basic behaviour. (a) Illustration of the probabilistic instrumental learning task. In each reward trial, the participants select one of the two options (the fractal images) and either receive a reward ("You win 1 point!") or nothing, depending on the probability assigned to the chosen option. In each avoidance trial, participants receive either a loss ("You lose 1 point!") or nothing. In each neutral trial, a visual image not associated with monetary incentives is presented with the message ("No change!"). ITI, intertrial interval. (b) The proportion of correct choices in reward and avoidance trials. Here, the correct choice is defined as selecting the option that provides a higher reward (or lower loss) probability in the given trial. HC, healthy control; OCD, obsessive-compulsive disorder; PG, pathological gambling. (c) Mean response time (seconds) in the reward and avoidance trials. $^{**}P < 0.01$, two-tailed Welch's *t* test, Bonferroni-corrected for the two tests performed: HC vs. OCD, and HC vs. PG. Summary data to reproduce the figure are available at https://osf.io/v7em5/.

https://doi.org/10.1371/journal.pbio.3002031.g001

To earn rewards and avoid losses as much as possible, they needed to learn the reward/loss probability of each stimulus through experience. In the neutral trials, a visual image not associated with reward or loss was presented, depending on the participants' choices. By computational modelling of behaviour, we examined whether behavioural effects of reward prediction errors (i.e., learning rates in RL) are altered in OCD and PG. Furthermore, we sought to identify neural underpinnings of RL anomalies, focusing on prediction error signals in the striatum, mPFC, ACC, and insula.

Results

Demographics and basic behaviour

There were no significant differences between the OCD/PG and HC groups in age (Welch's *t* test; HC versus OCD: $t = 0.900$ and $P_{corr} = 0.744$; HC versus PG: $t = 0.274$ and $P_{corr} = 1.000$), sex ratio (Fisher's exact test; HC versus OCD: $P_{corr} = 1.000$; HC versus PG: $P_{corr} = 1.000$), and IQ (Welch's *t* test; HC versus OCD: $t = 0.456$ and $P_{corr} = 1.000$; HC versus PG: $t = 0.522$ and $P_{corr} = 1.000$). As expected, the questionnaire-based OCD symptoms (i.e., the score on the Obsessive-Compulsive Inventory-Revised scale [53]) were more severe in the OCD group than in the HC and PG groups (Welch's *t* test; OCD versus HC: $t = 11.096$ and $P_{corr} < 0.001$; OCD versus PG: $t = 5.266$ and $P_{corr} < 0.001$). Moreover, the PG group exhibited problematic gambling (i.e., the score on the Problem Gambling Severity Index [54]) to a higher extent than the HC and OCD groups (Welch's *t* test; PG versus HC: $t = 7.853$ and $P_{corr} < 0.001$; PG versus OCD: $t = 7.677$ and $P_{corr} < 0.001$). Additional details are provided in S1 Table and S1 Fig.

Overall task performance did not differ between the OCD/PG and HC groups (Fig 1B). There were no significant differences in the proportion of correct choices (i.e., selection of the option with the higher/lower reward/loss probability) in the reward trials (Welch's *t* test; HC versus OCD: $t = 1.553$ and $P_{corr} = 0.251$; HC versus PG: $t = 2.197$ and $P_{corr} = 0.066$) or the

avoidance trials (Welch's t test; HC versus OCD: $t = 0.508$ and $P_{corr} = 1.000$; HC versus PG: $t = 1.065$ and $P_{corr} = 0.589$), while the proportion in the reward trials was significantly lower in the PG group than in the HC group without the multiple comparison correction ($P = 0.033$). Furthermore, we observed no significant differences in response time between the OCD/PG and HC groups, except that OCD patients had longer reaction time than HC in the avoidance trials (Fig 1C; Welch's t test; reward trial: HC versus OCD, $t = 1.484$ and $P_{corr} = 0.288$; HC versus PG, $t = 0.223$ and $P_{corr} = 1.000$; and avoidance trial: HC versus OCD, $t = 2.954$ and $P_{corr} = 0.009$; HC versus PG, $t = 0.840$ and $P_{corr} = 0.815$).

Behaviour: The effect of the past outcome and choice

We next analysed the trial-by-trial choice data to confirm that participants' behaviour was driven by past reward and loss outcomes. Furthermore, we also examined how past choices affected their behaviour [36]. Generalised linear mixed models (GLMM1) revealed that the past reward and loss outcomes had significant impacts on current choice behaviour (Fig 2). Consistent with the RL framework, past rewards in the reward trials had a significantly positive effect (Fig 2A; HC: $b = 1.204$, $t = 10.454$, and $P_{corr} < 0.001$; OCD: $b = 0.952$, $t = 7.409$, and $P_{corr} < 0.001$; PG: $b = 0.920$, $t = 5.977$, and $P_{corr} < 0.001$), and past losses in the avoidance trials had a significantly negative effect (Fig 2B; HC: $b = -1.290$, $t = 13.150$, and $P_{corr} < 0.001$; OCD: $b = -1.428$, $t = 12.832$, and $P_{corr} < 0.001$; PG: $b = -1.302$, $t = 9.002$, and $P_{corr} < 0.001$) across all groups. Past outcomes in the neutral trials had no effect (S2A Fig; HC: $b = -0.031$, $t = 0.290$, and $P_{corr} = 1.000$; OCD: $b = 0.010$, $t = 0.081$, and $P_{corr} = 1.000$. PG: $b = -0.119$, $t = 0.756$, and $P_{corr} = 0.900$), as expected given that they were not associated with any monetary incentives. Furthermore, in both the reward and avoidance trials, the effect of past choices was significantly positive (Fig 2; reward trial: HC, $b = 0.434$, $t = 5.884$, and $P_{corr} < 0.001$; OCD, $b = 0.371$, $t = 4.324$, and $P_{corr} < 0.001$; PG, $b = 0.243$, $t = 2.346$, and $P_{corr} = 0.038$; and avoidance trial: HC, $b = 0.922$, $t = 13.211$, and $P_{corr} < 0.001$; OCD, $b = 1.142$, $t = 13.929$, and $P_{corr} < 0.001$; PG, $b = 1.010$, $t = 10.023$, and $P_{corr} < 0.001$), indicating a tendency to repeat the same choice.

We also tested for group differences (HC versus OCD and HC versus PG) in the effects of past outcomes and choices on behaviour (GLMM2). The additional analyses revealed no differences between the groups in either the reward trials (HC versus OCD: $t = 1.559$ and $P_{corr} = 0.238$ for the effect of past outcomes, and $t = 0.245$ and $P_{corr} = 1.000$ for the effect of past choices; HC versus PG: $t = 1.532$ and $P_{corr} = 0.251$ for the effect of past outcomes, and $t = 1.360$ and $P_{corr} = 0.348$ for the effect of past choices) or the avoidance trials (HC versus OCD:

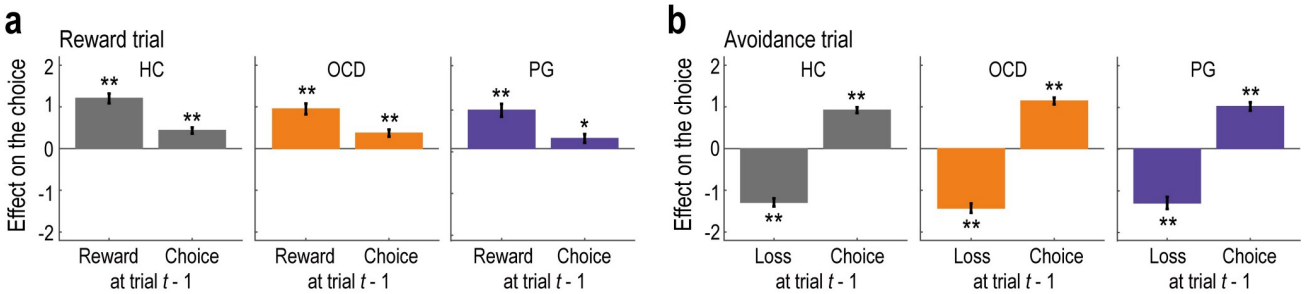

Fig 2. Model-neutral regression analysis on the behaviour. (a) Effects of past rewards and choices on current behaviour in reward trials (mean ± SEM, estimated using the three generalised linear mixed-effect models, GLMM1). $^{**}P < 0.01$ and $^{*}P < 0.05$, two-tailed Welch's t test, Bonferroni-corrected for the two tests performed in each model. HC, healthy control; OCD, obsessive-compulsive disorder; PG, pathological gambling. (b) Effects of past losses and choices on current behaviour in avoidance trials. The format is the same as in (a). Summary data to reproduce the figure are available at https://osf.io/v7em5/.

https://doi.org/10.1371/journal.pbio.3002031.g002

$t = 0.913$ and $P_{corr} = 0.723$ for the effect of past outcomes, and $t = 2.093$ and $P_{corr} = 0.073$ for the effect of past choices; HC versus PG: $t = 0.002$ and $P_{corr} = 1.000$ for the effect of past outcomes, and $t = 0.847$ and $P_{corr} = 0.794$ for the effect of past choices) (S2B and S2C Fig).

These results were maintained when different lengths of the outcome and choice history were considered in the GLMMs. For example, when the previous three trials were considered, the total effects of past rewards/losses (i.e., the sum of the regression coefficients over the past three trials: b_{sum}) were significantly positive/negative (reward trial: HC, $b_{sum} = 1.392$, $F = 49.479$, and $P_{corr} < 0.001$; OCD, $b_{sum} = 1.612$, $F = 53.293$, and $P_{corr} < 0.001$; PG, $b_{sum} = 1.284$, $F = 25.782$, and $P_{corr} < 0.001$; and avoidance trial: HC, $b_{sum} = -1.964$, $F = 131.317$, and $P_{corr} < 0.001$; OCD, $b_{sum} = -1.998$, $F = 105.714$, and $P_{corr} < 0.001$; PG, $b_{sum} = -2.114$, $F = 63.258$, and $P_{corr} < 0.001$). Furthermore, no significant differences were found between the groups in either the reward trials (HC versus OCD: $F = 0.245$ and $P_{corr} = 1.000$ for the effect of past outcomes, and $F = 0.068$ and $P_{corr} = 1.000$ for the effect of past choices; HC versus PG: $F = 0.259$ and $P_{corr} = 1.000$ for the effect of past outcomes, and $F = 4.155$ and $P_{corr} = 0.084$ for the effect of past choices) or the avoidance trials (HC versus OCD: $F = 0.007$ and $P_{corr} = 1.000$ for the effect of past outcomes, and $F = 1.383$ and $P_{corr} = 0.480$ for the effect of past choices; HC versus PG: $F = 0.160$ and $P_{corr} = 1.000$ for the effect of past outcomes, and $F = 0.424$ and $P_{corr} = 1.000$ for the effect of past choices).

Behaviour: Computational model fits

In theory, the above results can reflect non-trivial interactions among multiple components, such as value learning, perseveration, and stochastic noise in the decision-making process [42]. To disentangle these components, we fitted RL models to participants' choice data and analysed the best-fitted model (see Methods for details).

RL1 was a conventional RL model, in which the value of the chosen option was updated in proportion to the reward prediction error with learning rate α. The learning rate governed the extent to which new reward/loss information was incorporated into the value updating process. RL2 had different learning rates, $\alpha_{(+)}$ and $\alpha_{(-)}$, for positive and negative reward prediction errors, respectively. RL3 included the perseveration effect, in which the constant bonus, γ—denoting the degree of perseveration (i.e., the tendency to repeat the same choice)—was added to the option chosen in the previous trial. RL4 included both differential learning rates and perseveration. In the remainder of this section, we only present RL1, RL2, and RL3, as model recovery analysis on simulated data revealed that the complicated model (RL4) was not well dissociable from the simpler one (RL3) (see S1 Text for detail). For example, RL4 generated the simulated data; however, RL3 and RL4 were identified as the best-fit models with probabilities of 0.37 and 0.47, respectively (S4A Fig), demonstrating poor identifiability between the two models. We confirmed that the key findings (i.e., group differences in the learning rates) did not change whether RL4 is included or not (S1 Text and S4C and S4D Fig). In the model fitting, a hierarchical modelling approach was employed to reduce the estimation noise [55]. Each model's goodness of fit was assessed using the widely applicable Akaike information criterion (WAIC) [56]. We validated the procedure using parameter- and model-recovery analyses of the simulated data [57] (S3 Fig). We also confirmed that the best-fitted models could replicate the behavioural results obtained in the model-neutral regression analyses (S5 Fig; cf. Figs 1B and 2).

The model comparison revealed that in the reward trials, RL with asymmetric learning rates (RL2) provided the best fit for all three groups (HC, OCD, and PG; Fig 3A). This implies that in reward-seeking decision-making, participants recruited different systems for learning from outcomes that were better (positive prediction error) and worse (negative prediction

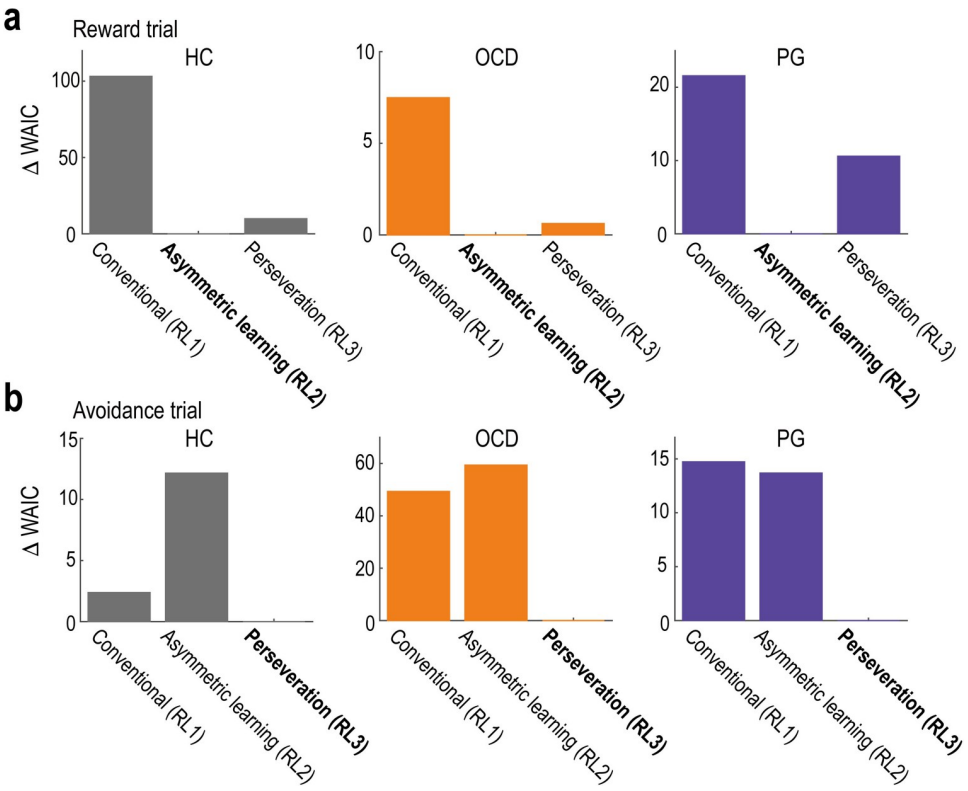

Fig 3. Computational model fitting to behaviour. (a) Model comparison in reward trials. We plotted each model's WAIC value relative to the best model (note: a smaller value indicates a better fit). The best-fit model is highlighted in bold. HC, healthy control; OCD, obsessive-compulsive disorder; PG, pathological gambling; RL, reinforcement learning; WAIC, widely applicable Akaike information criterion. (b) Model comparison in the avoidance trials. The format is the same as in (a). Summary data to reproduce the figure are available at https://osf.io/v7em5/.

https://doi.org/10.1371/journal.pbio.3002031.g003

error) than expected. In contrast, in the avoidance trials, RL with perseveration (RL3) provided the best fit for all groups (Fig 3B). This suggests that in loss-avoidance decision-making, participants employed a common unitary system for learning from positive and negative prediction errors and that they preferred the option that had been previously chosen regardless of its outcome.

In addition to the main RL models, we tested parsimonious models that share a common set of parameters across the two trial types. We also examined a model with motor-perseveration: i.e., the tendency to choose options presented on the same side (left or right) in consecutive trials irrespective of the trial types. Furthermore, we tested another model that involved a context-dependent adaptation of the outcome values [58,59], as well as models that endogenously modulated the learning rate [60,61]. These additional models, however, did not outperform the best fit RL models (see S1 Text for details).

Behaviour: Parameters of the best-fitted models

We further examined whether and how the key decision parameters in the best-fitted models differed between the OCD/PG and HC groups. In the reward trials, compared with the HC group, participants with PG exhibited a higher learning rate for positive reward prediction errors, $\alpha_{(+)}$, and a lower rate for negative prediction errors, $\alpha_{(-)}$ (Fig 4A and 4C; $\alpha_{(+)}$, Bayes factor (BF) = 167.71 and 95% highest density interval (HDI) = [0.12, 0.41]; $\alpha_{(-)}$, BF = 40.41 and 95% HDI = [−0.33, −0.09]). Compared to the HC group, we found a lower learning rate for

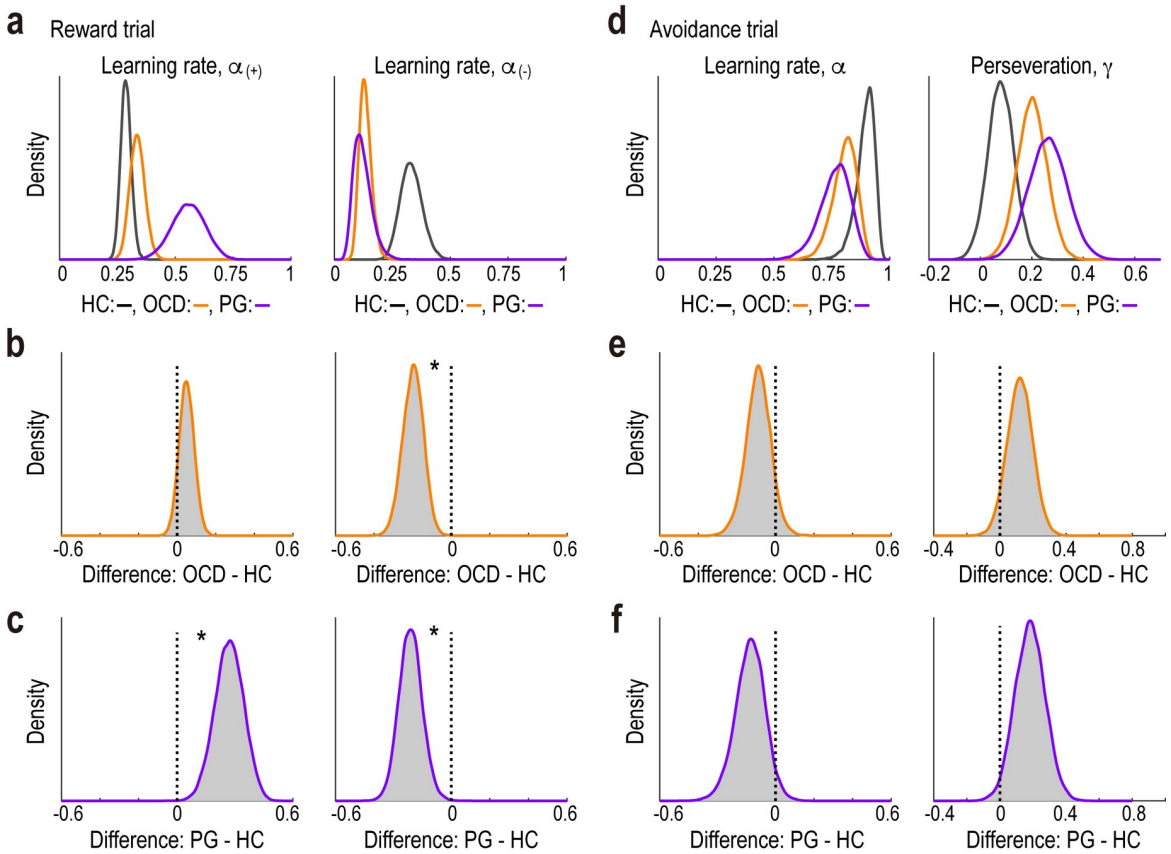

Fig 4. Parameter estimates in the best-fitted models. (a) Posterior distributions of group-level mean parameters of RL2 in reward trials for each group. *Left*, learning rate from the positive reward prediction error, $\alpha_{(+)}$. *Right*, learning rate from the negative reward prediction error, $\alpha_{(-)}$. *Grey*, healthy control (HC); *orange*, obsessive-compulsive disorder (OCD); *purple*, pathological gambling (PG). (b) Posterior distributions of the difference between the OCD and HC groups in reward trials. *Left*, $\alpha_{(+)}$; and *right*, $\alpha_{(-)}$. *BF > 20 denotes "very strong" or "strong" evidence [100] for the difference. BF, Bayes factor. (c) Posterior distributions of the difference between the PG and HC groups in reward trials. The format is the same as in (b). (d) Posterior distributions of group-level mean parameters of RL3 in avoidance trials for each group. *Left*: learning rate, α; *right*: perseveration, γ. The format is the same as in (a). (e) Posterior distributions of the difference between the OCD and HC groups in avoidance trials. *Left*, α; *right*, γ. The format is the same as in (b). (f) Posterior distributions of the difference between the PG and HC groups in avoidance trials. The format is the same as in (b). Summary data to reproduce the figure are available at https://osf.io/v7em5/.

https://doi.org/10.1371/journal.pbio.3002031.g004

negative prediction errors in the OCD group (Fig 4A and 4B; $\alpha_{(-)}$, BF = 104.90 and 95% HDI = [−0.31, −0.09]). In contrast, in the avoidance trials, there was no evidence for differences in key parameter values—learning rate (α) and perseveration (γ)—between the OCD/PG and HC groups (Fig 4D–4F; HC versus OCD: α, BF = 0.47 and 95% HDI = [−0.22, 0.04] and γ, BF = 0.06 and 95% HDI = [−0.03, 0.27]; and HC versus PG: α, BF = 0.97 and 95% HDI = [−0.29, 0.02] and γ, BF = 0.17 and 95% HDI = [0.01, 0.36]). Taken together, these results suggest that PG is characterised by excessive sensitivity (insensitivity) to positive (negative) prediction errors, respectively, whereas OCD is characterised by insensitivity to negative prediction errors in reward-seeking decision-making.

As a robustness check, we obtained the parameter estimates without assuming a hierarchical structure (i.e., each participant's individual-level parameters were drawn from group-specific higher-level distributions: see S3A Fig). First, assuming all the participants in each group share the same set of the parameter values (S6A Fig), we fitted the best-fitted models by pooling all the participants' data in each group. Next, assuming each participant has his/her own

(independent) set of parameters (S6B Fig), we fitted the models to each participant's data separately. These additional analyses showed results consistent with those in the original hierarchical model fitting (S6C–S6F Fig): i.e., higher/lower learning rates for positive/negative reward prediction errors in the PG group, and a lower rate for the negative prediction error in the OCD group.

We also examined the associations between questionnaire-based symptom severity and the decision parameters ($\alpha_{(+)}$, $\alpha_{(-)}$, and γ) within each of the three groups. However, we did not find any significant correlations for either reward or avoidance trials (S2 Table), suggesting that the decision parameters were not direct predictors of symptom severity.

Neuroimaging: Reward prediction error in the reward trials

Before proceeding with the main analyses of interest, we replicated previous well-established findings [62,63]. That is, in healthy people, reward prediction error and expected value signals are represented in the ventral striatum and mPFC. We fitted the RL models to data from the HC group to the (pooled) reward and avoidance trials and confirmed that the ventral striatum encodes the reward prediction error at the time of outcome delivery (S7A–S7C Fig; $P < 0.05$, cluster-level corrected). We also confirmed that the mPFC encodes the value of the chosen option at the time of decision-making (S7D–S7F Fig.; $P < 0.05$, cluster-level corrected).

In the main analysis, we employed a two-step approach. First, we identified the brain regions tracking the reward prediction error independent of diagnosis by averaging the data of all participants. Specifically, we obtained a weighted average, with weights equal to the inverse of the number of participants in each group (see Methods for details), which indicated the regions of interest (ROIs) that were statistically independent of the subsequent across-group comparisons. Next, we tested whether these neural representations differed between the OCD/PG and HC groups.

The results of the behavioural modelling suggest that in reward-seeking decision-making, learning from positive and negative reward prediction errors was governed by distinct systems (see Fig 5 and the legend for a schematic illustration and formal definition of positive and negative reward prediction errors). Therefore, we hypothesised that the positive and negative parts of the prediction error were processed separately in the brain in the reward trials.

Across the three groups in the reward trials, positive reward prediction errors were significantly correlated with the BOLD signal in the ventral striatum (ventral part of the caudate and putamen extending to the nucleus accumbens), mPFC (Brodmann area, BA 8/32/10), and insula at the time of outcome presentation (Fig 6A; $P < 0.05$, cluster-level corrected; see S3 Table for other activated areas). Comparisons between the HC and PG groups further demonstrated that the neural encoding of positive prediction error in the insula, but not that in the

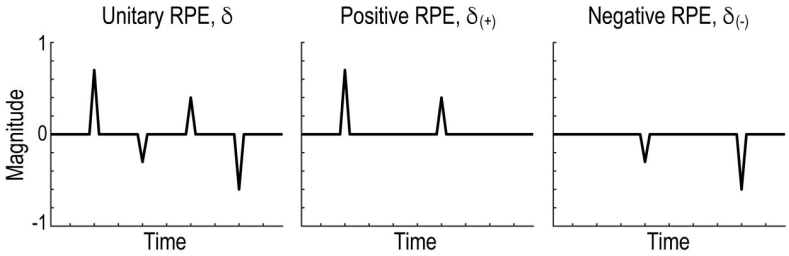

Fig 5. Schematic illustration of the three types of reward prediction error. *Left*, conventional unitary reward prediction error δ. *Centre*, positive reward prediction error, $\delta_{(+)} = max(\delta, 0)$. *Right*, negative reward prediction error, $\delta_{(-)} = min(\delta, 0)$. RPE, reward prediction error.

https://doi.org/10.1371/journal.pbio.3002031.g005

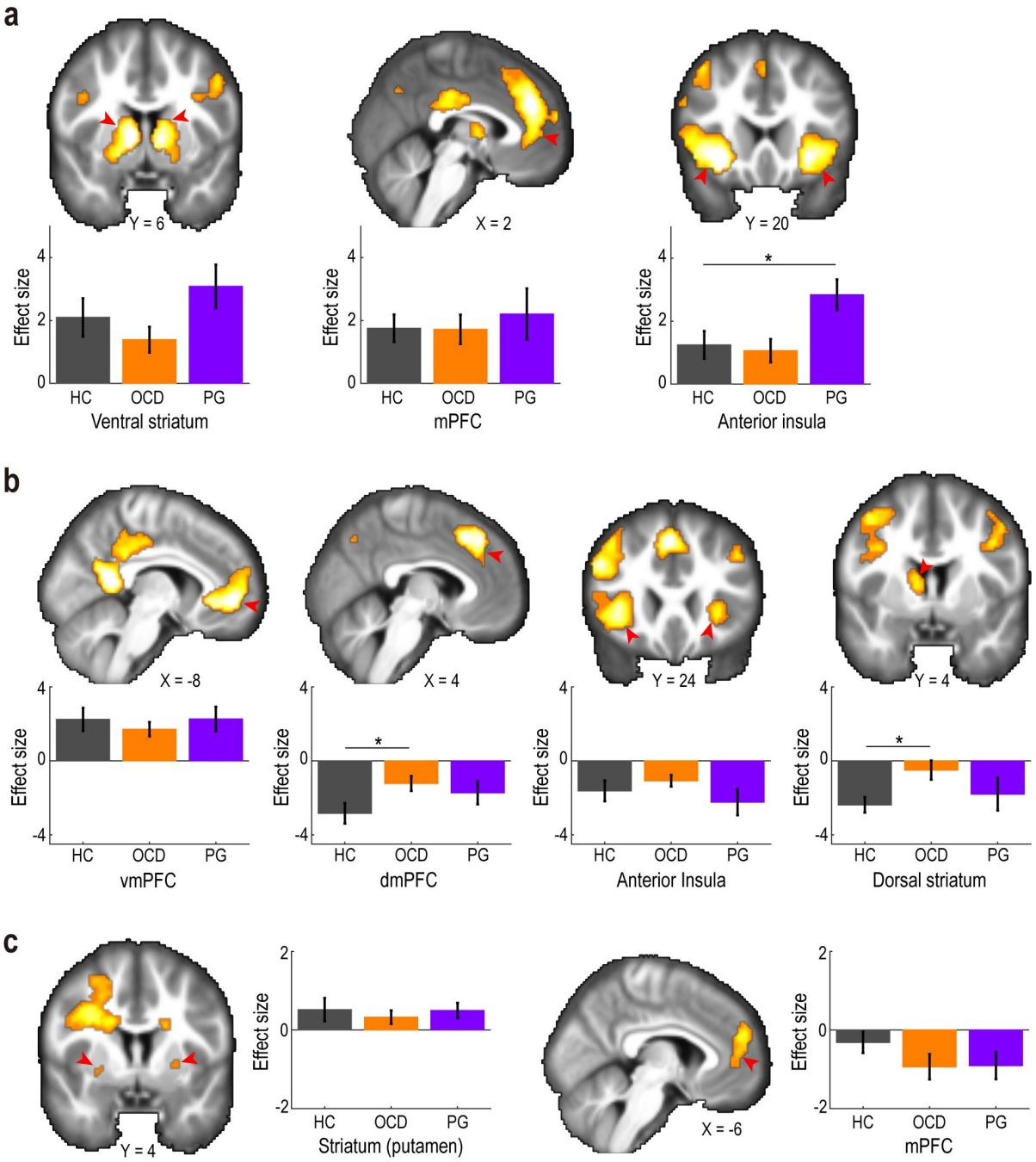

Fig 6. Neural correlates of reward prediction errors. (a) Positive reward prediction error in the reward trials. The activation maps are thresholded at $P < 0.0005$ (uncorrected) for display purposes. Red arrows indicate regions of interest. Bar-plots show effect sizes of the positive reward prediction error in the HC, OCD, and PG groups (mean ± SEM). *Grey*, healthy control (HC); *orange*, obsessive-compulsive disorder (OCD); *purple*, pathological gambling (PG). $^{*}P < 0.05$ in a two-tailed Welch's *t* test, Bonferroni-corrected for the two tests performed. mPFC, medial prefrontal cortex. (b) Negative prediction error in the reward trials. vmPFC, ventromedial prefrontal cortex; dmPFC, dorsomedial prefrontal cortex. The format is the same as in (a). (c) Unitary reward prediction error in the avoidance trials. The format is the same as in (a). Summary data to reproduce the figure are available at https://osf.io/v7em5/.

https://doi.org/10.1371/journal.pbio.3002031.g006

ventral striatum or mPFC, was greater in PG than in HC (Fig 6A and S8 Fig.; insula: $t = 2.430$ and $P_{corr} = 0.039$; ventral striatum: $t = 1.076$ and $P_{corr} = 0.577$; and mPFC: $t = 0.491$ and $P_{corr} = 1.000$). However, no significant difference was detected between the OCD and HC groups in any ROIs (Fig 6A; insula: $t = 0.324$ and $P_{corr} = 1.000$; ventral striatum: $t = 0.963$ and $P_{corr} = 0.680$; and mPFC: $t = 0.055$ and $P_{corr} = 1.000$). We also performed a whole-brain search to explore brain regions that differentially encoded positive prediction errors between the OCD/PG and HC participants. The whole-brain analysis found no regions under the statistical threshold ($P < 0.05$, cluster-level corrected).

Negative reward prediction errors in the reward trials were positively correlated with the BOLD signal in the ventromedial prefrontal cortex (vmPFC, BA 32/10) at the time of outcome presentation across the three groups (Fig 6B; $P < 0.05$, cluster-level corrected). Moreover, the negative prediction error was negatively correlated with the BOLD signal in the dorsomedial prefrontal cortex (dmPFC, BA 8), insula, and dorsal striatum (caudate) (Fig 6B, $P < 0.05$, cluster-level corrected) across the three groups (see S4 Table for a comprehensive list of the activated areas). A comparison of the HC and OCD groups revealed that the encoding of negative prediction errors in the dmPFC and the dorsal striatum was weaker in the OCD group (Fig 6B; dmPFC: $t = 2.359$ and $P_{corr} = 0.043$; and dorsal striatum: $t = 2.831$ and $P_{corr} = 0.013$; vmPFC: $t = 0.730$ and $P_{corr} = 0.938$; and insula: $t = 0.851$ and $P_{corr} = 0.798$). The comparison between HC and PG did not reveal significant differences in any of the four ROIs (dmPFC: $t = 1.330$ and $P_{corr} = 0.383$; and dorsal striatum: $t = 0.592$ and $P_{corr} = 1.000$; vmPFC: $t = 0.021$ and $P_{corr} = 1.000$; and insula: $t = 0.674$ and $P_{corr} = 1.000$). The whole-brain search also did not detect differential encoding of negative reward prediction errors.

Neuroimaging: Reward prediction error in the avoidance trials

Since RL with a unitary learning rate (RL3) best explained participants' behaviour in loss-avoidance decision-making (Fig 3B), we next aimed to identify brain regions encoding the unitary reward prediction error (Fig 5) in avoidance trials. The prediction error was found to be significantly correlated with BOLD signals in the striatum (putamen) and mPFC (BA 10/9) at the time of outcome in the avoidance trials across the three groups (Fig 6C; $P < 0.05$, cluster-level corrected; see S5 Table for other activated areas). The strength of the neural encodings in the striatum and mPFC did not differ significantly between the OCD/PG and HC groups (Fig 6C; striatum: HC versus OCD, $t = 0.563$ and $P_{corr} = 1.000$, and HC versus PG, $t = 0.058$ and $P_{corr} = 1.000$; and mPFC: HC versus OCD, $t = 1.478$ and $P_{corr} = 0.290$, and HC versus PG, $t = 1.351$ and $P_{corr} = 0.371$). The whole-brain search revealed no brain regions that differentially encoded reward prediction errors between OCD/PG and HC.

Neuroimaging: Post hoc analyses

In this experiment, approximately half of the OCD patients were medicated with selective serotonin reuptake inhibitors (SSRIs). We examined how the medication affects key findings in OCD, i.e., attenuated encoding of negative reward prediction errors in the dmPFC and striatum during reward trials (Fig 6B). The additional analysis revealed no significant differences in the encoding strengths of negative prediction errors between medicated and unmedicated OCD patients (dmPFC: $t = 0.359$ and $P = 0.722$; dorsal striatum: $t = 1.062$ and $P = 0.299$), while encoding strengths in medicated OCD patients were between those in unmedicated OCD patients and the HC group (S9 Fig).

A recent perspective [64] by Lebreton and colleagues alerted us to a pitfall in interpreting individual differences in the strength of neural encoding of a computational variable (e.g., reward prediction error): The precise interpretation depends on whether the range of neural

activity is adapted to the domain of the variable of interest (i.e., "range-adaptation" coding) or not (i.e., "proportional" coding). The authors claim that typical studies in computational psychiatry, such as our current study, have implicitly assumed proportional coding (i.e., native variables, rather than the standardised variables, are fed into the fMRI analysis), although the coding principle has been unexplored. In other words, the findings of those studies could be spurious if the brain employs a range-adaptation coding. Following the recommendation provided in the perspective [64], we presented the neuroimaging results with z-normalisation of the prediction errors alongside the original ones without normalisation (S10 Fig). The additional data analysis revealed that the effect sizes decreased by an average of 30% from the original ones, while the overall patterns remained unchanged. These findings caution against evaluating the original results of altered neural encodings of reward prediction errors in OCD/PG.

Discussion

Using an RL framework, this study both identified the existence of RL anomalies in, and elucidated basic neural computations that are altered in patients with OCD and PG. Compared with healthy controls, patients with OCD learned less from worse-than-expected experiences in reward-seeking decision-making, with attenuated representations of the negative reward prediction error in the dmPFC and dorsal striatum. Furthermore, patients with PG exhibited excessive (deficient) learning from better-than- (worse-than-) expected experiences, respectively, reflecting the enhanced representation of positive reward prediction error in the insula. On the other hand, we did not find any behavioural or neural signatures of OCD and PG in loss-avoidance decision-making.

Although behavioural inflexibility is thought to be a hallmark of OCD, the underlying computational mechanisms remain elusive [24,65] (but see [66]). Several studies have addressed this issue by incorporating a so-called perseveration or stickiness parameter, which controls for the tendency to repeat the same choice (regardless of the outcome) in the RL model [37,38]. However, neither of the previous studies nor the present one found any evidence for an increased perseveration parameter in patients with OCD (compared with healthy controls). Instead, a decreased perseveration has been reported in OCD [37,38,40]. In theory, inflexibility can also be captured by asymmetric learning for positive and negative reward prediction errors. That is, excessive learning from better-than-expected experiences and/or deficient learning from worse-than-expected experiences may result in the repetitive choice of the same option [42,43]. Consistent with the theoretical consideration, we demonstrated that, in reward-seeking decision-making, the learning rate for negative reward prediction errors was lower in patients with OCD than in healthy controls. This finding supports the notion of asymmetric learning, and may suggest a novel computational mechanism of behavioural inflexibility in OCD.

Past studies to date have demonstrated altered reward-learning in OCD [38,40,50,67], but the specific pattern of results has been mixed. Marzuki and colleagues [40] demonstrated that adolescents with OCD have lower and higher learning rates from negative and positive reward prediction errors, respectively. The decreased learning rate for negative prediction errors was consistent with our findings, but this study did not observe the increased rate for positive error. On the other hand, another study [38] by Kanen and colleagues reported the opposite pattern: an increased learning rate from negative prediction errors in OCD, contrasting our findings but consistent with the enhanced performance monitoring hypothesis that OCD is associated with increased sensitivity to negative feedback [24,68]. Furthermore, using a novel predictive inference task, Vaghi and colleagues [67] demonstrated an increased unitary learning rate that does not discriminate between positive or negative prediction errors in OCD.

We did not find evidence for altered perseveration in OCD, although previous studies reported a decreased perseveration [37,38,40]. This discrepancy may be due to the differences in tasks. In our task, three pairs of fractal images (corresponding to the three types of trials; see Fig 1A) were interleaved across trials in each session. In this setting, perseveration does not necessarily imply choosing the same option (image) in consecutive trials. In contrast, only one pair of images was presented within a session of the tasks used in the previous studies [37,38,40], where perseveration indicates choices of the same option in consecutive trials. This difference may reconcile the seemingly inconsistent findings. Furthermore, in the present study, we excluded participants who chose only one of the options in some sessions (i.e., never chose the other option), which could affect the null finding of perseveration in OCD.

We found that, in reward-seeking decision-making, the decreased learning rate from worse-than-expected outcomes in OCD was associated with altered neural encoding of negative reward prediction errors in both the dorsal striatum (caudate) and the dmPFC (including the ACC). Specifically, neural activity in those regions was negatively coupled with negative prediction error, and the couplings were weakened in OCD (Fig 6B). The dmPFC finding is broadly consistent with that of a previous study [37] by Hauser and colleagues. They showed an altered negative coupling between ACC activity and reward prediction errors in patients with OCD. The study found that neural activity in the ACC was negatively correlated with unitary reward prediction error independent of diagnosis. The negative correlation was attenuated in patients with OCD compared to healthy controls (i.e., the correlation was highly negative in the healthy controls but not in OCD patients). The pattern of the altered prediction error signals in ACC is partly consistent with what was observed in the dmPFC's negative prediction error signal (i.e., negative component of the unitary prediction error; see Fig 5). A subsequent study further demonstrated that the neural response to reward prediction errors in the ACC was modulated by a dopamine D2/3 receptor agonist, pramipexole, and a dopamine D2/3 receptor antagonist, amisulpride [50]. The caudate nucleus is a major projecting locus of dopamine neurons in the substantia nigra pars compacta [69] and a core region of the OCD-related neural network [70,71]. For example, reduced resting-state functional connectivity between the caudate and lateral prefrontal cortex is associated with the impairment of cognitive flexibility in patients with OCD [72]. Elevated activity in the caudate has also been implicated in the formation of excessive habits in patients with OCD [73]. Our neuroimaging results, together with the previous findings, suggest that the altered dopaminergic neural responses to reward prediction error in the dmPFC and striatum (caudate) underlie the deficient learning from worse-than-expected experiences in OCD.

Behavioural inflexibility in OCD has been assessed using various tasks [24] including probabilistic reversal learning [37,40], delayed feedback [74], and deterministic set-shifting tasks (e.g., Wisconsin Card Sorting task) [75]. Notably, Gillan and colleagues examined behavioural inflexibility in OCD concerning an imbalance between goal-directed and habitual behaviours [76,77] using tasks that involve two-stage decision-making, devaluation, or contingency degradation [78]. One caveat in this study is whether and how much our findings in the probabilistic instrumental learning tasks can be generalised to other contexts.

The computational modelling of behaviour in patients with PG revealed increased/decreased learning rates for positive/negative reward prediction errors for reward-seeking decision-making. In other words, PG patients learned more from better-than-expected experiences and less from worse-than-expected experiences than healthy controls, which may lead to addictive gambling despite the adverse consequences. The learning-based account has been indirectly suggested by previous studies on neural responses to monetary reward and loss in PG patients [46]. However, no study to date has demonstrated direct evidence. For example, previous studies of RL in PG did not explore asymmetric learning [39,41,79,80], while they employed reward-

learning tasks. To our knowledge, the present study is the first to provide direct evidence for this account by fitting a formal RL model with asymmetric learning rates to behaviour.

We found increased insula neural activity in response to positive reward prediction error in PG patients in reward-seeking decision-making. This result is consistent with previous findings that a brain network including the insula is associated with behaviours implicated in PG [81–83]. In PG patients, a craving for gambling is correlated with enhanced reactivity to the gambling cue (but not food cues) in the insula [83]. Furthermore, patients with damage to the insula exhibit less gambling-related cognitive distortions [82]. These findings implicate the insula in PG-related overlearning from positive information and may contribute to the development of novel treatments for PG that target neural activity in the insula.

We did not find behavioural or neural differences between OCD/PG patients and healthy controls in loss-avoidance decision-making. The null results could be due to the small effect size relative to the sample size. We used monetary loss as an aversive stimulus in the experiment, whereas previous studies used electrical shocks [73,84]. The difference in the psychological/neural intensity between the primary (electrical shocks) and secondary (monetary losses) punishments may cause a smaller effect size in our study. More research is needed to further examine avoidance decision-making in OCD and PG.

In this study, participants with OCD and PG performed the same decision-making task in the same environment. This experimental design enabled the elucidation of similarities and differences between OCD and PG. Despite the similarity in behavioural signatures (e.g., deficits in flexibility), the DSM-5 classifies PG (gambling disorder) as an addictive disorder that is separate from OCD and impulse-control disorders [27,30,33,34]. Our findings are consistent with the possibility that compulsive behaviours observed in the two disorders are driven by a similar computational process (i.e., asymmetric reward learning) but implemented by distinct neural mechanisms (the dmPFC and caudate in OCD and the insula in PG).

Caveats and limitations of the study

One caveat of this study is its real-world significance. There was no significant difference in basic behaviour (i.e., the overall task performance and reaction time) between the OCD/PG and HC groups. Furthermore, we found no significant correlations between decision parameters (i.e., learning rate and perseveration) and questionnaire-based measurements of symptom severity within each group. Moreover, it is worth noting that there was no direct evidence of behavioural inflexibility in our sample of OCD and PG patients. We did not collect explicit measures of behavioural inflexibility, and generalised linear regression did not reveal any significant differences in the effect of past choices on behaviour between the OCD/PG and HC groups. The relationship between asymmetric learning rates obtained from our RL modelling and behavioural inflexibility observed in the real world (e.g., excessive hand washing and pathological gambling) remains elusive. Taken together, future studies should be conducted to evaluate the real-world significance of this task.

A limitation of this study is the small sample size, which may result in low statistical power. Our study's sample size (34 HC, 29 OCD patients, and 17 PG patients) was comparable to those of other neuroimaging studies that recruited patients with OCD/PG [37–39,50,73,80,83]. However, it was smaller than the sample size recommended for exploring individual differences [85]. Therefore, caution must be exercised when interpreting and applying this study's findings because low statistical power could increase the likelihood of obtaining false-negative and false-positive results [86].

In this study, the two clinical groups were recruited from different channels: OCD from clinical services and PG from the local community (see Methods). Furthermore, the

proportion of the participants with the intake of SSRIs was higher in OCD than in PG (S1 Table, Fisher's exact test; OCD versus PG: $P = 0.023$). These factors could lead to the differential behavioural and neural signatures of RL [87–89].

Conclusions

Grounded in a well-characterised RL framework, the present study elucidates neurocomputational mechanisms of reward-learning anomalies in patients with OCD and PG. We demonstrate that OCD and PG were associated with altered learning rates from positive and negative reward prediction errors. Our findings are consistent with the possibility that asymmetric sensitivity to better- and worse-than-expected outcomes in reward-learning regulates, at least in part, abnormal compulsive behaviours in these disorders. More broadly, the present study sheds light on the importance of a computational model that bridges behaviour and neural activity in a unified framework (e.g., RL), facilitating more comprehensive explanations of the pathophysiology of mental disorders.

Methods

This study was approved by the Human Research Ethics Committee of Monash University (ID: 1238896).

Participants

We recruited 34 participants diagnosed with OCD, 23 with PG, and 39 HC as part of a broader project [32,90,91]. No statistical methods were used to predetermine the sample size; however, our sample size was consistent with those used in previous studies [37,39,50,73,83]. All participants provided informed written consent.

Participants with OCD were recruited from specialist clinical services located in Melbourne, VIC, Australia. Participants with PG and HCs were recruited from the local community. The participants were pre-assessed to exclude those with a lifetime history of concussion, neurological disease, or drug abuse/dependence.

All the participants underwent phone screening, followed by an in-depth face-to-face assessment. The first screening over the phone included the short MINI International neuropsychiatric interview screener, the Florida Obsessive-Compulsive Inventory [92], and the Problem Gambling Severity Index [54]. In the second screening, LP and LB performed the full MINI International neuropsychiatric interview for OCD and other mental disorders and the structured clinical interview for PG in DSM-IV to further characterise the participants' symptoms. Participants with OCD were screened to ensure that their severity section score on the Florida obsessive-compulsive inventory was >8, and their diagnosis was confirmed using treatment services and the MINI International neuropsychiatric interview. Participants with PG who engaged in electronic gambling at least once a week were screened to ensure that their problem gambling severity index score was >8, and their diagnosis was confirmed by the Structured Clinical Interview for DSM-IV. Participants with OCD or PG who had either depression or anxiety (as indexed by the MINI) were not excluded as long as the OCD and PG symptoms constituted the primary cause of distress and interference in the participants' lives. Participants were excluded if they met the criteria for any other psychiatric disorder, including the concurrent presence of OCD and PG.

Furthermore, on the day of the second screening, we obtained the Obsessive-Compulsive Inventory-Revised (OCI-R) [53], Barratt Impulsiveness Scale [93], Beck Depression Inventory [94], and State and Trait Anxiety Inventory [95] scores as indices of the severity of psychiatric symptoms.

Exclusion criteria

For the analyses, we excluded data from participants who failed to respond to >20% of the trials in any of the four main conditions (reward and avoidance trials conducted in sessions 1 and 2). We also excluded those who chose only one option in more than one of the four conditions. The remaining data used for the subsequent analyses included 34 HC (18 females; age 34.6 ± 9.8 years; 33 right-handed), 29 OCD (16 females; age 32.3 ± 10.1; 29 right-handed), and 17 PG (7 females; age 33.6 ± 12.9; 17 right-handed) participants. See S1 Table for the demographic and clinical information.

Stimuli

We used six pairs of 12 fractal images in the experiment. Each pair was presented to the participants in one of the three types of trials in each of the two sessions. The association between image pairs and trial types was randomised across participants.

Decision-making task

Each participant performed a probabilistic instrumental learning task (Fig 1A). In this task, each of two sessions contained randomly interleaved 30 *reward*, 30 *avoidance*, and 30 *neutral* trials (i.e., one of the three trials was presented at pseudo-random on each trial). Participants chose between two stimuli to earn rewards in the reward trials and made choices to avoid losses in the avoidance trials. In the neutral trials, a visual image not associated with monetary incentives was presented. Note that different pairs of stimuli were presented in the three types of trial (see Fig 1A).

In the reward/avoidance trials, one of the paired stimuli yielded a reward/loss with a probability of approximately 0.7, whereas the other yielded a reward/loss with a probability of approximately 0.3. In the neutral trials, the two stimuli led to a neutral feedback with probabilities of approximately 0.7 and 0.3, respectively. In each session, the stimulus with a higher probability of reward/loss was randomly reversed once between the 10th and 20th trial without any explicit cue (the timing was independent across the three types of trials). The reversal was introduced to foster participants' learning over the course of all trials. Note that the reversal often favours model-based learning algorithms with an adaptive rate [96], but such models did not outperform the original models in this study (S1 Text). Furthermore, unlike a conventional reversal-learning task in which one of the two options was rewarding in each trial, in our task, the probabilistic outcomes of the two stimuli were determined independently. The participants did not receive any instructions regarding the structure of the reward/loss environment.

In each trial, participants chose between the two fractal images by pressing a button with their right hand within 2 s (decision phase; Fig 1A). The two images were positioned randomly to the left or right of the screen in each trial. After the response, the chosen image was highlighted by an increase in brightness (anticipation phase, 3.5 s). The outcome of the choice was then revealed by visual feedback (outcome phase, 1.5 s; Fig 1A): a picture of a store card with the text "You win 1 point!" indicated a reward; a red cross on top of the card with the text "You lose 1 point!" indicated a loss; a scrambled picture of the card with the text "Nothing" indicated a neutral outcome; and a blank screen with a fixation cross indicated no reward and no loss. Note that the neutral outcome was not associated with any reward or loss and had essentially the same meaning as the blank screen in terms of monetary incentives.

Reward payment

The participants were informed that they would receive a gift card corresponding to the number of points earned in the tasks. However, all the participants received a $20 gift card that

could be used in an Australian major department store (Myer) and a supermarket (Coles). The deception was employed in the experimental design due to ethical concerns regarding paying PG patients based on their performance. However, it is worth noting that we replicated well-known striatal and prefrontal responses to the reward prediction error and the expected value signals (S7 Fig) [62,63], supporting the validity of our experiment to examine reward-based decision-making.

Behavioural data analysis

GLMM1. To examine the effects of the past outcome and choice on behaviour, we constructed a generalised linear mixed-effects model for each of the 3 trial types and the 3 groups (i.e., 9 models in total, called GLMM1 collectively): logit P (choice = A) ~ 1 + $R_{t\text{-}1}$ + $C_{t\text{-}1}$ + (1 | participant), where $R_{t\text{-}1}$ and $C_{t\text{-}1}$ denote the outcome and choice in the previous trial (i.e., the previous occasion where the same trial type was presented) [36]. $R_{t\text{-}1}$ was coded as 1 if the participant chose option A and obtained a reward (loss) in the previous trial, -1 if they chose option B and obtained a reward (loss), and 0 if there was no reward (loss). $C_{t\text{-}1}$ was coded as 1 if the participant chose option A in the previous trial and -1 otherwise. The term (. | participants) indicates that the variables were considered as random effects (i.e., were allowed to vary between participants). We also tested models that included the random effects of $R_{t\text{-}1}$ and $C_{t\text{-}1}$; however, the complicated models were not favoured by the Akaike information criterion (AIC).

GLMM2. We performed additional GLMM analyses to test for group differences (HC versus OCD and HC versus PG) in the effects of the past outcome and choice: logit P (choice at t = A) ~ 1 + $R_{t\text{-}1}$ + $C_{t\text{-}1}$ + d_{OCD} x $R_{t\text{-}1}$ + d_{OCD} x $C_{t\text{-}1}$ + d_{PG} x $R_{t\text{-}1}$ + d_{PG} x $C_{t\text{-}1}$ + d_{OCD} + d_{PG} + (1 | participant), where d_{OCD} and d_{PG} are dummy variables indicating the OCD and PG groups. Here, the interaction terms denote how the effects of the past reward and choice were modulated by mental disorders (compared with HC).

Correction for multiple comparisons. To control for the type I error rate in the behavioural analyses, we employed Bonferroni correction for the number of tests performed (see each of the figure legends for details).

Computational models

We considered four RL models. RL1 was a conventional one, in which the value of the chosen option was updated in proportion to the reward prediction error with learning rate α. RL2 had different learning rates, $\alpha_{(+)}$ and $\alpha_{(-)}$, for positive and negative reward prediction errors, respectively. RL3 included a perseveration parameter, γ, denoting the tendency to repeat the same choice. RL4 included both differential learning rates and perseveration.

RL1. In this simple Q-learning model [9], an individual's decision was guided by the values of the two available options, $Q(A)$ and $Q(B)$. These values governed the participant's probability of choosing option A as follows: $q(A) = 1 / [1 + \exp(-\beta (Q(A)-Q(B)))]$, where the parameter $\beta>0$ (*inverse temperature*) controls for the degree of noise in the decision-making. Once the outcome of the choice is revealed, the value of the chosen option is updated by the reward prediction error. That is, when option A is chosen, the value is updated as follows: $Q(A) \leftarrow Q(A)+\alpha\,\delta$. Here, δ is the reward prediction error, defined as $\delta = R-Q(A)$, where R denotes the reward outcome (1 for a reward, -1 for a loss, and 0 for no-reward or no-loss). The parameter $\alpha \in [0,1]$ is called the *learning rate* and controls the speed of learning with respect to the reward prediction error.

RL2. This model included differential learning rates for positive and negative reward prediction errors [97,98]. Specifically, the value of the chosen option, say A, was updated as

follows:

$$Q(A) \leftarrow \begin{cases} Q(A) + \alpha_{(+)}\delta \text{ if } \delta \geq 0 \\ Q(A) + \alpha_{(-)}\delta \text{ if } \delta < 0 \end{cases},$$

where $\alpha_{(+)}$ and $\alpha_{(-)}$ indicate the asymmetric learning rates.

RL3. The third model included the perseveration effect, in which the constant bonus γ was added to the option chosen in the previous trial (i.e., the previous occasion where the same trial type was presented) [37,38]. Specifically, the probability of choosing option A was given by $q(A) = 1 / [1 + \exp(-\beta(Q(A)-Q(B) + \gamma)]$ if A had been chosen in the previous trial, and by $1 / [1 + \exp(-\beta(Q(A)-Q(B))-\gamma)]$ when B had been chosen. The parameter γ denotes the degree of perseveration (i.e., tendency to repeat the same choice). The process of value updating (i.e., learning) was the same as that in RL1.

RL4. The fourth model included differential learning rates and perseveration term.

These models were fitted to the participants' choice data in each group (HC, OCD, and PG) and each trial type (reward and avoidance trials). For this, we used a hierarchical modelling approach to reduce the estimation noise in the parameter estimates [55] (see S3A Fig for the details of RL3). Each participant's individual-level parameters were assumed to be drawn from common group-level normal distributions: e.g., for participant i, the parameters were defined as $\text{logit}(\alpha^{(i)}) \sim N(\mu_\alpha, \sigma_\alpha)$, $\log(\beta^{(i)}) \sim N(\mu_\beta, \sigma_\beta)$, and $\gamma^{(i)} \sim N(\mu_\gamma, \sigma_\gamma)$. The posterior distributions of the group-level parameters were estimated using the variational Bayes framework in *CmdStanR* version 2.26.1, as sampling methods (e.g., Markov chain Monte Carlo) were not practically feasible for exhaustive model- and parameter-recovery analyses (see *Validation of the model fitting based on the simulation data* below). Weekly informative priors were set for the group-level parameters: $N(0,3)$ for μ_α and μ_γ, and $N(0,1)$ for μ_β, σ_α, σ_β, and σ_γ. We confirmed that the main results of the model fitting (Figs 3 and 4) did not change by ensuring flatter priors for the group-level parameters: $N(0,5)$ for μ_α and μ_γ, and $N(0,2)$ for μ_β, σ_α, σ_β, and σ_γ; or $N(0,9)$ for μ_α and μ_γ, and $N(0,3)$ for μ_β, σ_α, σ_β, and σ_γ.

Each model's goodness of fit was assessed using the WAIC [56], obtained from the posterior distributions using the *R* package "*loo*" (version 2.4.1). The WAIC is an advanced version of AIC, which is applicable to a wide range of cases (e.g., hierarchical modelling).

To examine group differences in the estimated parameters, we focused on the posterior distributions of the mean differences (Fig 4B, 4C, 4E, and 4F). We calculated the Bayes Factor (BF: the strength of evidence against the null hypothesis that the group difference is *zero*) using the Savage–Dickey method [99]. Conventionally, BF >150 is judged as "very strong," 20 to 150 as "strong," and 3 to 20 as "positive evidence" [100].

Validation of the model fitting based on the simulation data

We ran the simulation 300 times. In each simulation run, the data of 30 agents were generated by each of the three computational models (i.e., RL1, RL2, and RL3). To cover the ranges of the parameters to a reasonable degree, we assumed that α, $\alpha_{(+)}$, and $\alpha_{(-)}$ were sampled from $Beta(1, 1)$, that β was from $N(3.5, 0.5)$, and that γ was sampled from $N(0.4, 0.2)$.

Parameter recovery analysis. In the simulation data, we first examined whether the generative parameter values could be recovered by hierarchical Bayesian model fitting [57]. For each parameter in each model, we first plotted the fitted values against the true generative values (S3B–S3D Fig, *top*), and then quantified the recoverability by the correlation between the two values (S3B–S3D Fig, *bottom*; histograms of the correlation coefficients over 300 simulation runs).

Model recovery analysis. We next verified the identifiability of the models by constructing a "confusion matrix" and an "inversion matrix" [57] (S3E Fig). In the confusion matrix, each row denoted the probability of each of the three competing models providing the best fit (i.e., with the lowest WAIC value) for the data generated by the corresponding model. In the inversion matrix, each row denoted the probability that the data best fitted by the corresponding model (i.e., with the lowest WAIC value) was generated from each of the three competing models.

Replication of the results of model-neutral analysis. The data simulated by the best-fitted models and their parameter estimates successfully replicated the main behavioural findings obtained in the model-neutral analysis (S5 Fig).

fMRI data acquisition

We collected fMRI images using a 3T Siemens MAGNETOM Skyra Syngo MR D13C scanner at Monash Biomedical Imaging (Clayton, VIC, Australia). The BOLD signal was measured using a one-shot T2*-weighted echo planar imaging sequence (volume TR = 2,000 ms, TE = 30 ms, FA = 90˚). We acquired 34 oblique slices (thickness, 3.0 mm; gap, 0 mm; FOV, 230 × 230 mm; matrix, 76 × 76) per volume. After the two functional runs (479 volumes each), high-resolution (1 mm^3) anatomical images were acquired using a standard MPRAGE pulse sequence (TR = 2,300 ms, TE = 2.07 ms, FA = 9˚). The data were analysed using the SPM12 software in MATLAB R2020b on a MacBook Pro (16-inch, 2019; Mac OS X 10.17.7). Data from one OCD participant were excluded due to technical problems with the fMRI scan.

fMRI data analysis

We analysed the fMRI data to identify the neural correlates of model-derived reward prediction errors in the reward and avoidance trials (compared to the counterparts in the neutral trials). The comparison with the data from neutral trials was important to control for confounding effects, such as visual responses to the outcome feedback [101]. The counterparts of the reward prediction error in the reward and avoidance trials were the original and inverted prediction errors in the neutral trials, respectively. Note that in the reward trials, reward prediction error was heightened with the unexpected presentation of a visual feedback of reward and lowered with the unexpected omission of the visual feedback. In the avoidance trials, reward prediction error was heightened with the unexpected omission of a visual feedback of loss and lowered with the unexpected presentation of the visual feedback. In the neutral trials, the prediction error was heightened with the unexpected presentation of a visual feedback of the neutral outcome and lowered with the unexpected omission of the visual feedback. Given the nature of the prediction error in each trial type, we compared the neural encoding of prediction error in the reward trials to that of the prediction error in the neutral trials, and the neural encoding of prediction error in the avoidance trials to that of the inverse prediction error in the neutral trials.

Preprocessing. We employed a standard procedure in SPM12: After slice timing correction, the images were realigned to the first volume to correct for the participants' movements, spatially normalised based on the segmentation of the anatomical image, and spatially smoothed using an 8 mm FWHM Gaussian kernel. High-pass temporal filtering (filter width = 128 s) was also applied to the data.

GLM1. The GLM contained parametric regressors representing the unitary reward prediction errors at the time of the outcome (Figs 1A and 5). Specifically, the participant-specific design matrices contained the following regressors: two boxcar functions for the decision (duration = response time) and outcome (duration = 1.5 s) phases, and one stick function at the time of the key response (duration = 0 s). Missed trials were modelled as separate

regressors. We included three parametric modulators of the boxcar function for the outcome phase that encoded the reward prediction errors in the reward, avoidance, and neutral trials. To control for the nuisance effects of decision-related neural signals, we included three modulators in the decision phase that encoded the value of the chosen option in the reward, avoidance, and neutral trials. All regressors were convolved with a canonical hemodynamic response function (with the serial orthogonalization of parametric modulators turned off). In addition, six motion-correction parameters were included as regressors of no interest to account for motion-related artefacts.

GLM2. This GLM contained separate parametric regressors representing the positive and negative parts of the reward prediction error (Figs 1A and 5). That is, the design matrices contained six parametric modulators of the outcome phase: (1) the positive prediction error in the reward trials; (2) the negative prediction error in the reward trials; (3) the positive prediction error in the avoidance trials; (4) the negative prediction error in the avoidance trials; (5) the positive prediction error in the neutral trials; and (6) the negative prediction error in the neutral trials. The other settings were the same as those used for GLM1.

Whole-brain analysis. As described in the main text, we used a two-step approach. First, we identified the brain regions tracking the reward prediction error independent of the mental disorder diagnosis by averaging the data of all the participants. We obtained the weighted average using the inverse of the number of participants in each group as weights, which indicated the ROIs that were statistically independent of the subsequent across-group comparisons. To this end, we defined contrasts of interest for the prediction errors weighted by the inverse of the number of participants in each group (HC: $N = 34$, OCD: $N = 29$, and PG: $N = 17$) as follows: 17/34 for HC, 17/29 for OCD, and 1 for PG following normalisation by the number of PG patients. For each participant, the contrasts were estimated at every voxel of the whole brain and entered into a random-effects analysis. We set our significance threshold at $P < 0.05$, which was whole-brain corrected for multiple comparisons at the cluster level (based on the conservative cluster-forming threshold at $P < 0.0005$, uncorrected [102]).

ROI analysis. Second, we tested whether the neural representations differed between the OCD/PG and HC groups. ROI analyses were performed using the MarsBaR toolbox for SPM. Each ROI was defined as a sphere (radius: 9 mm = 3 voxels) centred on the peak coordinates in a cluster of interest (S3–S5 Tables). Group differences in the effect sizes of the neural representations (i.e., regression coefficients in the fMRI GLM) were examined using the two-tailed Welch's *t* test with Bonferroni correction for the number of tests performed (see each of the figure legends for details).

Supporting information

S1 Text. Supplementary note 1 (model fitting).
(PDF)

S1 Fig. Questionnaire-based psychiatric symptoms.
(PDF)

S2 Fig. Supplementary regression analysis on the behaviour.
(PDF)

S3 Fig. Supplementary behavioural analysis on the simulated data.
(PDF)

S4 Fig. Supplementary results of the model fitting with RL4.
(PDF)

S5 Fig. Supplementary behavioural analysis on the simulated data.
(PDF)

S6 Fig. Supplementary analysis of the parameter estimation.
(PDF)

S7 Fig. Replication of the previous neuroimaging findings in HC.
(PDF)

S8 Fig. Neuroimaging analysis for the left and right insula.
(PDF)

S9 Fig. Comparisons between SSRI-medicated and unmedicated OCD patients.
(PDF)

S10 Fig. Comparisons between the neuroimaging results with and without z-normalisation of reward prediction errors.
(PDF)

S1 Table. Demographic characteristics of the participants.
(PDF)

S2 Table. Correlations between symptom severity and RL parameters.
(PDF)

S3 Table. Brain areas exhibiting significant changes in the BOLD signal associated with the positive reward prediction error in reward trials.
(PDF)

S4 Table. Brain areas exhibiting significant changes in the BOLD signal associated with the negative reward prediction error in reward trials.
(PDF)

S5 Table. Brain areas exhibiting significant changes in the BOLD signal associated with the unitary reward prediction error in avoidance trials.
(PDF)

S6 Table. WAIC values in the supplementary model comparison.
(PDF)

S7 Table. WAIC values in the supplementary model comparison.
(PDF)

Acknowledgments

We thank Dr. Kentaro Katahira for the helpful discussion of the revised manuscript.

Author Contributions

Conceptualization: Shinsuke Suzuki, Xiaoliu Zhang, Amir Dezfouli, Leah Braganza, Ben D. Fulcher, Linden Parkes, Leonardo F. Fontenelle, Ben J. Harrison, Carsten Murawski, Murat Yücel, Chao Suo.

Data curation: Xiaoliu Zhang, Amir Dezfouli, Leah Braganza, Ben D. Fulcher, Linden Parkes, Leonardo F. Fontenelle, Ben J. Harrison, Carsten Murawski, Murat Yücel, Chao Suo.

Formal analysis: Shinsuke Suzuki, Amir Dezfouli, Ben D. Fulcher.

Funding acquisition: Murat Yücel.

Investigation: Shinsuke Suzuki, Xiaoliu Zhang, Amir Dezfouli, Leah Braganza, Ben D. Fulcher, Linden Parkes, Leonardo F. Fontenelle, Ben J. Harrison, Carsten Murawski, Murat Yücel, Chao Suo.

Methodology: Shinsuke Suzuki, Amir Dezfouli, Leah Braganza, Ben D. Fulcher, Linden Parkes, Leonardo F. Fontenelle, Ben J. Harrison, Carsten Murawski, Murat Yücel, Chao Suo.

Project administration: Xiaoliu Zhang, Leah Braganza, Ben D. Fulcher, Linden Parkes, Leonardo F. Fontenelle, Ben J. Harrison, Carsten Murawski, Murat Yücel, Chao Suo.

Supervision: Leonardo F. Fontenelle, Carsten Murawski, Murat Yücel, Chao Suo.

Visualization: Shinsuke Suzuki.

Writing – original draft: Shinsuke Suzuki.

Writing – review & editing: Shinsuke Suzuki, Xiaoliu Zhang, Amir Dezfouli, Leah Braganza, Ben D. Fulcher, Linden Parkes, Leonardo F. Fontenelle, Ben J. Harrison, Carsten Murawski, Murat Yücel, Chao Suo.

References

1. Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016; 19:404–413. https://doi.org/10.1038/nn.4238 PMID: 26906507
2. Rutledge RB, Chekroud AM, Huys QJ. Machine learning and big data in psychiatry: toward clinical applications. Curr Opin Neurobiol. 2019; 55:152–159. https://doi.org/10.1016/j.conb.2019.02.006 PMID: 30999271
3. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends Cogn Sci. 2012; 16:72–80. https://doi.org/10.1016/j.tics.2011.11.018 PMID: 22177032
4. Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. Lancet Psychiatry. 2014; 1:148–158. https://doi.org/10.1016/S2215-0366(14)70275-5 PMID: 26360579
5. Robinson OJ, Chase HW. Learning and Choice in Mood Disorders: Searching for the Computational Parameters of Anhedonia. Comput Psychiatry. 2017; 1:208–233. https://doi.org/10.1162/CPSY_a_00009 PMID: 29400358
6. Huys QJ, Pizzagalli DA, Bogdan R, Dayan P. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. Biology Mood Anxiety Disord. 2013; 3:12. https://doi.org/10.1186/2045-5380-3-12 PMID: 23782813
7. Powers AR, Mathys C, Corlett PR. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science. 2017; 357:596–600. https://doi.org/10.1126/science.aan3458 PMID: 28798131
8. Dayan P, Daw ND. Decision theory, reinforcement learning, and the brain. Cognitive Affect Behav Neurosci. 2008; 8:429–453. https://doi.org/10.3758/cabn.8.4.429 PMID: 19033240
9. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT Press; 1998.
10. Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. Science. 1997; 275:1593–1599. https://doi.org/10.1126/science.275.5306.1593 PMID: 9054347
11. Glimcher PW. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proc Natl Acad Sci U S A. 2011; 108:15647–15654. https://doi.org/10.1073/pnas.1014269108 PMID: 21389268
12. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science. 2004; 304:452–454. https://doi.org/10.1126/science.1094285 PMID: 15087550
13. Rutledge RB, Dean M, Caplin A, Glimcher PW. Testing the Reward Prediction Error Hypothesis with an Axiomatic Model. J Neurosci. 2010; 30:13525–13536. https://doi.org/10.1523/JNEUROSCI.1747-10.2010 PMID: 20926678

14. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS. Associative learning of social value. Nature. 2008; 456:245. https://doi.org/10.1038/nature07538 PMID: 19005555

15. Suzuki S, Harasawa N, Ueno K, Gardner JL, Ichinohe N, Haruno M, et al. Learning to Simulate Others' Decisions. Neuron. 2012; 74:1125–1137. https://doi.org/10.1016/j.neuron.2012.04.030 PMID: 22726841

16. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature. 2006; 442:1042–1045. https://doi.org/10.1038/nature05051 PMID: 16929307

17. Seymour B O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, et al. Temporal difference models describe higher-order learning in humans. Nature. 2004; 429:664–667. https://doi.org/10.1038/nature02581 PMID: 15190354

18. Corlett PR, Mollick JA, Kober H. Meta-analysis of human prediction error for incentives, perception, cognition, and action. Neuropsychopharmacology. 2022:1–11. https://doi.org/10.1038/s41386-021-01264-3 PMID: 35017672

19. Suzuki S, Yamashita Y, Katahira K. Psychiatric symptoms influence reward-seeking and loss-avoidance decision-making through common and distinct computational processes. Psychiat Clin Neuros. 2021. https://doi.org/10.1111/pcn.13279 PMID: 34151477

20. Mukherjee D, Filipowicz ALS, Vo K, Satterthwaite TD, Kable JW. Reward and punishment reversal-learning in major depressive disorder. J Abnorm Psychol. 2020; 129:810–823. https://doi.org/10.1037/abn0000641 PMID: 33001663

21. DeGuzman M, Shott ME, Yang TT, Riederer J, Frank GKW. Association of Elevated Reward Prediction Error Response With Weight Gain in Adolescent Anorexia Nervosa. Am J Psychiatry. 2017; 174:557–565. https://doi.org/10.1176/appi.ajp.2016.16060671 PMID: 28231717

22. Brown VM, Zhu L, Solway A, Wang JM, McCurry KL, King-Casas B, et al. Reinforcement Learning Disruptions in Individuals With Depression and Sensitivity to Symptom Change Following Cognitive Behavioral Therapy. Jama Psychiatry. 2021:78. https://doi.org/10.1001/jamapsychiatry.2021.1844 PMID: 34319349

23. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. Elife. 2016; 5:e11305. https://doi.org/10.7554/elife.11305 PMID: 26928075

24. Robbins TW, Vaghi MM, Banca P. Obsessive-Compulsive Disorder: Puzzles and Prospects. Neuron. 2019; 102:27–47. https://doi.org/10.1016/j.neuron.2019.01.046 PMID: 30946823

25. Boog M, Höppener P, VD Wetering BJM, Goudriaan AE, Boog MC, Franken IHA. Cognitive Inflexibility in Gamblers is Primarily Present in Reward-Related Decision Making. Front Hum Neurosci. 2014; 8:569. https://doi.org/10.3389/fnhum.2014.00569 PMID: 25165438

26. van Timmeren T, Daams JG, van Holst RJ, Goudriaan AE. Compulsivity-related neurocognitive performance deficits in gambling disorder: A systematic review and meta-analysis. Neurosci Biobehav Rev. 2018; 84:204–217. https://doi.org/10.1016/j.neubiorev.2017.11.022 PMID: 29203423

27. American-Psychiatric-Association. Diagnostic and Statistical Manual of Mental Disorders (DSM5). American Psychiatric Pub; 2013.

28. Goudriaan AE, Oosterlaan J, Beurs ED, Brink WVD. Neurocognitive functions in pathological gambling: a comparison with alcohol dependence, Tourette syndrome and normal controls. Addiction. 2006; 101:534–547. https://doi.org/10.1111/j.1360-0443.2006.01380.x PMID: 16548933

29. van Holst RJ, van den Brink W, Veltman DJ, Goudriaan AE. Why gamblers fail to win: A review of cognitive and neuroimaging findings in pathological gambling. Neurosci Biobehav Rev. 2010; 34:87–107. https://doi.org/10.1016/j.neubiorev.2009.07.007 PMID: 19632269

30. Fontenelle LF, Oostermeijer S, Harrison BJ, Pantelis C, Yücel M. Obsessive-Compulsive Disorder, Impulse Control Disorders and Drug Addiction. Drugs. 2011; 71:827–840. https://doi.org/10.2165/11591790-000000000-00000 PMID: 21568361

31. Figee M, Pattij T, Willuhn I, Luigjes J, van den Brink W, Goudriaan A, et al. Compulsivity in obsessive–compulsive disorder and addictions. Eur Neuropsychopharm. 2016; 26:856–868. https://doi.org/10.1016/j.euroneuro.2015.12.003 PMID: 26774279

32. Parkes L, Tiego J, Aquino K, Braganza L, Chamberlain SR, Fontenelle LF, et al. Transdiagnostic variations in impulsivity and compulsivity in obsessive-compulsive disorder and gambling disorder correlate with effective connectivity in cortical-striatal-thalamic-cortical circuits. Neuroimage. 2019; 202:116070. https://doi.org/10.1016/j.neuroimage.2019.116070 PMID: 31382045

33. Potenza MN, Koran LM, Pallanti S. The relationship between impulse-control disorders and obsessive–compulsive disorder: A current understanding and future research directions. Psychiat Res. 2009; 170:22–31. https://doi.org/10.1016/j.psychres.2008.06.036 PMID: 19811840

34. Potenza MN. Should addictive disorders include non-substance-related conditions? Addiction. 2006; 101:142–151. https://doi.org/10.1111/j.1360-0443.2006.01591.x PMID: 16930171

35. Gershman SJ, Pesaran B, Daw ND. Human Reinforcement Learning Subdivides Structured Action Spaces by Learning Effector-Specific Values. J Neurosci. 2009; 29:13524–13531. https://doi.org/10.1523/jneurosci.2469-09.2009 PMID: 19864565

36. Lau B, Glimcher PW. DYNAMIC RESPONSE-BY-RESPONSE MODELS OF MATCHING BEHAVIOR IN RHESUS MONKEYS. J Exp Anal Behav. 2005; 84:555–579. https://doi.org/10.1901/jeab.2005.110-04 PMID: 16596980

37. Hauser TU, Iannaccone R, Dolan RJ, Ball J, Hättenschwiler J, Drechsler R, et al. Increased fronto-striatal reward prediction errors moderate decision making in obsessive–compulsive disorder. Psychol Med. 2017; 47:1246–1258. https://doi.org/10.1017/s0033291716003305 PMID: 28065182

38. Kanen JW, Ersche KD, Fineberg NA, Robbins TW, Cardinal RN. Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. Psychopharmacology. 2019; 236:2337–2358. https://doi.org/10.1007/s00213-019-05325-w PMID: 31324936

39. Wiehler A, Chakroun K, Peters J. Attenuated Directed Exploration during Reinforcement Learning in Gambling Disorder. J Neurosci. 2021; 41:2512–2522. https://doi.org/10.1523/jneurosci.1607-20.2021 PMID: 33531415

40. Marzuki AA, Tomić I, Ip SHY, Gottwald J, Kanen JW, Kaser M, et al. Association of Environmental Uncertainty With Altered Decision-making and Learning Mechanisms in Youths With Obsessive-Compulsive Disorder. Jama Netw Open. 2021; 4:e2136195. https://doi.org/10.1001/jamanetworkopen.2021.36195 PMID: 34842925

41. Carlisi CO, Norman L, Murphy CM, Christakou A, Chantiluke K, Giampietro V, et al. Shared and Disorder-Specific Neurocomputational Mechanisms of Decision-Making in Autism Spectrum Disorder and Obsessive-Compulsive Disorder. Cereb Cortex. 2017; 27:5804–5816. https://doi.org/10.1093/cercor/bhx265 PMID: 29045575

42. Katahira K. The statistical structures of reinforcement learning with asymmetric value updates. J Math Psychol. 2018; 87:31–45. https://doi.org/10.1016/j.jmp.2018.09.002

43. Palminteri S. Choice-confirmation bias and gradual perseveration in human reinforcement learning. PsyArXiv. 2021. https://doi.org/10.31234/osf.io/dpqj6

44. Figee M, Vink M, de Geus F, Vulink N, Veltman DJ, Westenberg H, et al. Dysfunctional Reward Circuitry in Obsessive-Compulsive Disorder. Biol Psychiatry. 2011; 69:867–874. https://doi.org/10.1016/j.biopsych.2010.12.003 PMID: 21272861

45. Kaufmann C, Beucke JC, Preuße F, Endrass T, Schlagenhauf F, Heinz A, et al. Medial prefrontal brain activation to anticipated reward and loss in obsessive–compulsive disorder. Neuroimage Clin. 2013; 2:212–220. https://doi.org/10.1016/j.nicl.2013.01.005 PMID: 24179774

46. Clark L, Boileau I, Zack M. Neuroimaging of reward mechanisms in Gambling disorder: an integrative review. Mol Psychiatry. 2019; 24:674–693. https://doi.org/10.1038/s41380-018-0230-2 PMID: 30214041

47. Balodis IM, Kober H, Worhunsky PD, Stevens MC, Pearlson GD, Potenza MN. Diminished Frontostriatal Activity During Processing of Monetary Rewards and Losses in Pathological Gambling. Biol Psychiatry. 2012; 71:749–757. https://doi.org/10.1016/j.biopsych.2012.01.006 PMID: 22336565

48. Simona R, Maria C, Luigi T, Gabriella S. The Neural Basis of Gambling Disorder: An Activation Likelihood Estimation Meta-Analysis. Neurosci Biobehav Rev. 2020; 120:279–302. https://doi.org/10.1016/j.neubiorev.2020.11.027 PMID: 33275954

49. Leyton M, Vezina P. On Cue: Striatal Ups and Downs in Addictions. Biol Psychiatry. 2012; 72:e21–e22. https://doi.org/10.1016/j.biopsych.2012.04.036 PMID: 22789688

50. Murray GK, Knolle F, Ersche KD, Craig KJ, Abbott S, Shabbir SS, et al. Dopaminergic drug treatment remediates exaggerated cingulate prediction error responses in obsessive-compulsive disorder. Psychopharmacology. 2019; 236:2325–2336. https://doi.org/10.1007/s00213-019-05292-2 PMID: 31201476

51. Kayser A. Dopamine and Gambling Disorder: Prospects for Personalized Treatment. Curr Addict Rep. 2019; 6:65–74. https://doi.org/10.1007/s40429-019-00240-8 PMID: 31728265

52. van Holst RJ, Sescousse G, Janssen LK, Janssen M, Berry AS, Jagust WJ, et al. Increased Striatal Dopamine Synthesis Capacity in Gambling Addiction. Biol Psychiatry. 2018; 83:1036–1043. https://doi.org/10.1016/j.biopsych.2017.06.010 PMID: 28728675

53. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, et al. The Obsessive-Compulsive Inventory: Development and Validation of a Short Version. Psychol Assess. 2002; 14:485–496. https://doi.org/10.1037/1040-3590.14.4.485

54. Ferris J, Wynne H. The Canadian Problem Gambling Index: Final Report. 2001.

55. Daw ND. Trial-by-trial data analysis using computational models. In: Delgado MR, Phelps Elizabeth A, Robbins Trevor W, editors. Oxford Scholarship Online; 2011.

56. Watanabe S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. Arxiv. 2010.

57. Wilson RC, Collins AG. Ten simple rules for the computational modeling of behavioral data. Elife. 2019; 8:e49547. https://doi.org/10.7554/elife.49547 PMID: 31769410

58. Burke CJ, Baddeley M, Tobler PN, Schultz W. Partial Adaptation of Obtained and Observed Value Signals Preserves Information about Gains and Losses. J Neurosci. 2016; 36:10016–10025. https://doi.org/10.1523/jneurosci.0487-16.2016 PMID: 27683899

59. Palminteri S, Khamassi M, Joffily M, Coricelli G. Contextual modulation of value signals in reward and punishment learning. Nat Commun. 2015; 6:8096. https://doi.org/10.1038/ncomms9096 PMID: 26302782

60. Piray P, Daw ND. A simple model for learning in volatile environments. PLoS Comput Biol. 2020; 16: e1007963. https://doi.org/10.1371/journal.pcbi.1007963 PMID: 32609755

61. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND. Differential roles of human striatum and amygdala in associative learning. Nat Neurosci. 2011; 14:1250–1252. https://doi.org/10.1038/nn.2904 PMID: 21909088

62. Daw ND O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441:876. https://doi.org/10.1038/nature04766 PMID: 16778890

63. O'Doherty JP, Hampton A, Kim H. Model-Based fMRI and Its Application to Reward Learning and Decision Making. Ann N Y Acad Sci. 2007; 1104:35–53. https://doi.org/10.1196/annals.1390.022 PMID: 17416921

64. Lebreton M, Bavard S, Daunizeau J, Palminteri S. Assessing inter-individual differences with task-related functional neuroimaging. Nat Hum Behav. 2019:1–9. https://doi.org/10.1038/s41562-019-0681-8 PMID: 31451737

65. Skvortsova V, Hauser T. Distinct computational mechanisms underlying cognitive flexibility deficits in impulsivity and compulsivity. 2022. https://doi.org/10.21203/rs.3.rs-1280535/v1

66. Fradkin I, Strauss AY, Pereg M, Huppert JD. Rigidly Applied Rules? Revisiting Inflexibility in Obsessive Compulsive Disorder Using Multilevel Meta-Analysis. Clin Psychological Sci. 2018; 6:481–505. https://doi.org/10.1177/2167702618756069

67. Vaghi MM, Luyckx F, Sule A, Fineberg NA, Robbins TW, Martino BD. Compulsivity Reveals a Novel Dissociation between Action and Confidence. Neuron. 2017; 96:348–354.e4. https://doi.org/10.1016/j.neuron.2017.09.006 PMID: 28965997

68. Gehring WJ, Himle J, Nisenson LG. Action-Monitoring Dysfunction in Obsessive-Compulsive Disorder. Psychol Sci. 1998; 11:1–6. https://doi.org/10.1111/1467-9280.00206 PMID: 11228836

69. Haber SN. The place of dopamine in the cortico-basal ganglia circuit. Neuroscience. 2014; 282:248–257. https://doi.org/10.1016/j.neuroscience.2014.10.008 PMID: 25445194

70. Baxter LR, Phelps ME, Mazziotta JC, Guze BH, Schwartz JM, Selin CE. Local Cerebral Glucose Metabolic Rates in Obsessive-Compulsive Disorder: A Comparison With Rates in Unipolar Depression and in Normal Controls. Arch Gen Psychiat. 1987; 44:211–218. https://doi.org/10.1001/archpsyc.1987.01800150017003 PMID: 3493749

71. Saxena S, Rauch SL. Functional neuroimaging and the neuroanatomy of obsessive-compulsive disorder. Psychiat Clin N Am. 2000; 23:563–586. https://doi.org/10.1016/s0193-953x(05)70181-7 PMID: 10986728

72. Vaghi MM, Vértes PE, Kitzbichler MG, Apergis-Schoute AM, van der Flier FE, Fineberg NA, et al. Specific Frontostriatal Circuits for Impaired Cognitive Flexibility and Goal-Directed Planning in Obsessive-Compulsive Disorder: Evidence From Resting-State Functional Connectivity. Biol Psychiat. 2017; 81:708–717. https://doi.org/10.1016/j.biopsych.2016.08.009 PMID: 27769568

73. Gillan CM, Apergis-Schoute AM, Morein-Zamir S, Urcelay GP, Sule A, Fineberg NA, et al. Functional Neuroimaging of Avoidance Habits in Obsessive-Compulsive Disorder. Am J Psychiatry. 2015; 172:284–293. https://doi.org/10.1176/appi.ajp.2014.14040525 PMID: 25526600

74. Sakai Y, Sakai Y, Abe Y, Narumoto J, Tanaka SC. Memory trace imbalance in reinforcement and punishment systems can reinforce implicit choices leading to obsessive-compulsive behavior. Cell Rep. 2022; 40:111275. https://doi.org/10.1016/j.celrep.2022.111275 PMID: 36044850

75. Chamberlain SR, Fineberg NA, Menzies LA, Blackwell AD, Bullmore ET, Robbins TW, et al. Impaired Cognitive Flexibility and Motor Inhibition in Unaffected First-Degree Relatives of Patients With Obsessive-Compulsive Disorder. Am J Psychiatry. 2007; 164:335–338. https://doi.org/10.1176/ajp.2007.164.2.335 PMID: 17267798

76. Gillan CM, Robbins TW. Goal-directed learning and obsessive-compulsive disorder. Philos Trans R Soc Lond B Biol Sci. 2014; 369:20130475–20130475. https://doi.org/10.1098/rstb.2013.0475 PMID: 25267818

77. Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, Enander J, et al. Disorders of compulsivity: a common bias towards learning habits. Mol Psychiatry. 2015; 20:345–352. https://doi.org/10.1038/mp.2014.44 PMID: 24840709

78. Balleine BW O'Doherty JP. Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology. 2010; 35:48–69. https://doi.org/10.1038/npp.2009.131 PMID: 19776734

79. Lei W, Liu K, Chen G, Tolomeo S, Liu C, Peng Z, et al. Blunted reward prediction error signals in internet gaming disorder. Psychol Med. 2020:1–10. https://doi.org/10.1017/s003329172000402x PMID: 33143778

80. Verdejo-Garcia A, Clark L, Verdejo-Román J, Albein-Urios N, Martinez-Gonzalez JM, Gutierrez B, et al. Neural substrates of cognitive flexibility in cocaine and gambling addictions. Br J Psychiatry. 2015; 207:158–164. https://doi.org/10.1192/bjp.bp.114.152223 PMID: 26045346

81. Potenza MN, Balodis IM, Derevensky J, Grant JE, Petry NM, Verdejo-Garcia A, et al. Gambling disorder. Nat Rev Dis Primers. 2019; 5:51. https://doi.org/10.1038/s41572-019-0099-7 PMID: 31346179

82. Clark L, Studer B, Bruss J, Tranel D, Bechara A. Damage to insula abolishes cognitive distortions during simulated gambling. Proc Natl Acad Sci U S A. 2014; 111:6098–6103. https://doi.org/10.1073/pnas.1322295111 PMID: 24711387

83. Limbrick-Oldfield EH, Mick I, Cocks RE, McGonigle J, Sharman SP, Goldstone AP, et al. Neural substrates of cue reactivity and craving in gambling disorder. Transl Psychiatry. 2017; 7:e992–e992. https://doi.org/10.1038/tp.2016.256 PMID: 28045460

84. Gillan CM, Morein-Zamir S, Urcelay GP, Sule A, Voon V, Apergis-Schoute AM, et al. Enhanced Avoidance Habits in Obsessive-Compulsive Disorder. Biol Psychiatry. 2014; 75:631–638. https://doi.org/10.1016/j.biopsych.2013.02.002 PMID: 23510580

85. Dubois J, Adolphs R. Building a Science of Individual Differences from fMRI. Trends Cogn Sci. 2016; 20:425–443. https://doi.org/10.1016/j.tics.2016.03.014 PMID: 27138646

86. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013; 14:365–376. https://doi.org/10.1038/nrn3475 PMID: 23571845

87. Palminteri S, Clair A-H, Mallet L, Pessiglione M. Similar Improvement of Reward and Punishment Learning by Serotonin Reuptake Inhibitors in Obsessive-Compulsive Disorder. Biol Psychiatry. 2012; 72:244–250. https://doi.org/10.1016/j.biopsych.2011.12.028 PMID: 22325972

88. McCabe C, Mishor Z, Cowen PJ, Harmer CJ. Diminished Neural Processing of Aversive and Rewarding Stimuli During Selective Serotonin Reuptake Inhibitor Treatment. Biol Psychiatry. 2010; 67:439–445. https://doi.org/10.1016/j.biopsych.2009.11.001 PMID: 20034615

89. Grossman CD, Bari BA, Cohen JY. Serotonin neurons modulate learning rate through uncertainty. Curr Biol. 2021. https://doi.org/10.1016/j.cub.2021.12.006 PMID: 34936883

90. Maleki S, Chye Y, Zhang X, Parkes L, Chamberlain SR, Fontenelle LF, et al. Neural correlates of symptom severity in obsessive-compulsive disorder using magnetization transfer and diffusion tensor imaging. Psychiatry Res Neuroimaging. 2020; 298:111046. https://doi.org/10.1016/j.pscychresns.2020.111046 PMID: 32106018

91. Zhang X, Chye Y, Braganza L, Fontenelle LF, Harrison BJ, Parkes L, et al. Severity related neuroanatomical alteration across symptom dimensions in obsessive-compulsive disorder. J Affect Disord Reports. 2021; 4:100129. https://doi.org/10.1016/j.jadr.2021.100129

92. Storch EA, Kaufman DAS, Bagner D, Merlo LJ, Shapira NA, Geffken GR, et al. Florida obsessive-compulsive inventory: Development, reliability, and validity. J Clin Psychol. 2007; 63:851–859. https://doi.org/10.1002/jclp.20382 PMID: 17674398

93. Patton JH, Stanford MS, Barratt ES. Factor structure of the barratt impulsiveness scale. J Clin Psychol. 1995; 51:768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1 PMID: 8778124

94. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An Inventory for Measuring Depression. Arch Gen Psychiatry. 1961; 4:561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004 PMID: 13688369

95. D. S C, L. G R, E. L R. Manual for the State-Trait Anxiety Inventory (Self-Evaluation Questionnaire). Consulting Psychologists Press; 1970.

96. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. Nat Neurosci. 2007; 10:1214–1221. https://doi.org/10.1038/nn1954 PMID: 17676057

97. Niv Y, Edlund JA, Dayan P, O'Doherty JP. Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. J Neurosci. 2012; 32:551–562. https://doi.org/10.1523/jneurosci.5498-10.2012 PMID: 22238090

98. Lefebvre G, Lebreton M, Meyniel F, Bourgeois-Gironde S, Palminteri S. Behavioural and neural characterization of optimistic reinforcement learning. Nat Hum Behav. 2017; 1:0067. https://doi.org/10.1038/s41562-017-0067

99. Wagenmakers E-J, Lodewyckx T, Kuriyal H, Grasman R. Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. Cogn Psychol. 2010; 60:158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001 PMID: 20064637

100. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc. 1995; 90:773–795. https://doi.org/10.1080/01621459.1995.10476572

101. Kim H, Shimojo S, O'Doherty JP. Is Avoiding an Aversive Outcome Rewarding? Neural Substrates of Avoidance Learning in the Human Brain. PLoS Biol. 2006; 4:e233. https://doi.org/10.1371/journal.pbio.0040233 PMID: 16802856

102. Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci U S A. 2016; 113:7900–7905. https://doi.org/10.1073/pnas.1602413113 PMID: 27357684