Computational modeling of reinforcement learning and functional neuroimaging of probabilistic reversal dissociates compulsive behaviors in Gambling and Cocaine Use Disorders

Katharina Zühlsdorff [1,2,3], Juan Verdejo-Román [4,5], Luke Clark [6], Natalia Albein-Urios [7], Carles Soriano-Mas [8,9,10], Rudolf N. Cardinal [2,11,13], Trevor W. Robbins [1,2], Jeffrey W. Dalley [1,2,11], Antonio Verdejo-García [12,*], Jonathan W. Kanen [1,2,*]

[1] Department of Psychology, University of Cambridge, Cambridge, UK

[2] Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK

[3] The Alan Turing Institute, London, UK

[4] Department of Personality, Assessment and Psychological Treatment, Universidad de Granada, Granada, Spain

[5] Mind, Brain and Behavior Research Center, Universidad de Granada, Granada, Spain

[6] Department of Psychology and Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC, Canada

[7] Cognitive Neuroscience Unit, School of Psychology, Deakin University, Melbourne, Australia

[8] Department of Psychiatry, Bellvitge Biomedical Research Institute-IDIBELL, Barcelona, Spain

[9] Department of Social Psychology and Quantitative Psychology, University of Barcelona, Spain

[10] CIBERSAM, Carlos III Health Institute, Spain

[11] Department of Psychiatry, University of Cambridge, Cambridge, UK

[12] School of Psychological Sciences and Turner Institute for Brain and Mental Health, Monash University, Melbourne, Australia

[13] Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK

[*]Joint last authors

Corresponding author: Katharina Zühlsdorff (kz294@cam.ac.uk), Department of Psychology, University of Cambridge, Downing Place, CB2 3EB, Cambridge, UK

Key words: cocaine use disorder, gambling disorder, reinforcement learning, prediction error, expected value, fMRI

Abstract

Background

Individuals with both Cocaine Use Disorder (CUD) and Gambling Disorder (GD) demonstrate impairments in cognitive flexibility (CF), the essential ability to adapt to changes in the environment. CF is commonly assessed in a laboratory setting using probabilistic reversal learning (PRL), which involves reinforcement learning (RL), the process by which feedback from the environment is used to adjust behavior.

Aims

It is poorly understood whether impairments in CF differ between individuals with CUD and GD, and how this is instantiated by the brain. We applied computational modelling of RL to gain a deeper mechanistic explanation of the latent processes underlying CF across two disorders of compulsivity.

Methods

We present a re-analysis of PRL data from individuals with either GD (n=18) or CUD (n=20), as well as control participants (n=18), using a hierarchical Bayesian RL approach. Furthermore, we relate behavioral findings to their underlying neural substrates through an analysis of task-based fMRI data.

Results

We observed lower 'stimulus stickiness' in GD. We also report differences in tracking expected values (EV) in individuals with GD compared to controls, with greater activity during reward EV tracking in the cingulate gyrus and amygdala. In CUD, we observed lower responses to positive punishment prediction errors (PPE) and greater activity following negative PPEs in the superior frontal gyrus compared to controls.

Conclusions

Using a computational approach to RL and CF, we show that individuals with GD and CUD differed in their perseverative tendencies and in how they tracked value neurally, which has implications for psychiatric classification.

Introduction

The diagnostic criteria for both Substance Use Disorders (SUD) and Gambling Disorder (GD) in the Diagnostic and Statistical Manual of Mental Disorders (fifth edition) (DSM-5) include unsuccessful attempts to stop substance abuse or gambling, jeopardizing relationships and educational/career opportunities, and financial troubles arising as a consequence of the disorder (1). Compulsivity, a key feature of both GD and SUDs, is defined as persistent actions inappropriate to a given situation, which have no clear relationship to the overall goal and frequently result in undesirable consequences (2). GD and SUDs are disorders of compulsivity and their behavioral phenotypes may thus overlap, but also diverge in certain aspects (3). Gaining a clearer definition of these phenotypes could inform the development of new treatments for disorders of compulsivity.

A further common feature of GD and SUD is behavioral inflexibility, defined as a deficit in adjusting behavior based on changes in environmental feedback (4,5). Individuals with a stimulant-related SUD exhibit higher rates of perseverative responding following a contingency change during probabilistic reversal learning (PRL), a paradigm used to investigate cognitive flexibility (6). During this task, individuals learn which action is associated with reward through trial and error. Following changes in stimulus contingencies, individuals need to flexibly adjust behavior. Indeed, reversal learning is impaired in rats and monkeys following prolonged exposure to cocaine (7,8).

Patients with GD, in comparison, show difficulties in learning novel stimulus-outcome associations following contingency changes during reversal learning (4). Following repeated negative feedback, patients with GD tend to stay rather than switch their response, or switch prematurely after little or no negative feedback during PRL (5). Individuals with GD perform significantly worse than healthy controls (HCs) on the Intra-/Extra-Dimensional Set Shifting test (IED), which assays higher order cognitive flexibility, with impairments observed at the extra-dimensional shift stage (requiring the most flexibility) (9). In a meta-analysis of participants diagnosed with GD on the related Wisconsin Card Sorting Test (WCST), patients made more perseverative errors than HCs (10). Overall, it is evident that individuals with GD are impaired on cognitive flexibility tasks and have greater perseverative tendencies, similar to individuals with SUD.

Reinforcement learning (RL) is the process by which positive and negative feedback from the environment is used to adjust behavior to maximize rewards and minimize punishment (11). In recent years, RL models have been used increasingly to gain deeper insights into the latent mechanisms underlying PRL on a trial-by-trial basis, which are represented by model parameters. One example of such a parameter is the exploration versus exploitation parameter, which reflects the extent to which learned values contribute to choice behavior. 'Stickiness' parameters track the tendency to repeatedly choose the same stimulus regardless of outcome (i.e., 'stimulus stickiness') or the tendency to repeat choices in the same location as before, irrespective of outcome (i.e., 'side stickiness'). These stickiness parameters fractionate the construct of perseveration as they parse different types of repetitive behaviors, for example, towards a location or a stimulus. Additionally, standard measures of perseveration assess behaviour following a contingency reversal, whereas stickiness accounts for a tendency to

repeat behaviors across all trials. Reward and punishment learning rates can also be determined via RL models, which index the speed at which the expected value of a choice is updated after a better than or worse than expected outcome (reward or punishment prediction error). Indeed, RL impairments following drug use and withdrawal have been demonstrated in rodents and humans. In rats, increased exploitation and stickiness have been reported after cocaine self-administration (12). Humans with SUD have also been found to have higher levels of stickiness, alongside greater punishment learning rates and lower reward learning rates (13). Critically, the RL fingerprint during PRL in GD has not been elucidated.

CUD has been associated with altered reward processing linked to differences in fronto-striatal activity. For example, a study employing functional magnetic resonance imaging (fMRI) has found that individuals diagnosed with CUD exhibited lower blood-oxygen level dependent (BOLD) signals in the orbitofrontal cortex (OFC) than control participants following monetary gains on a forced-choice task containing three monetary value conditions (14). Neural activity is also known to be altered in patients with SUD during PRL, such as in the middle frontal gyrus (MFG) and caudate nucleus, areas known to contribute to performance on this task (6,15). A meta-analysis of 52 studies reported that the OFC is hypoactive following detoxification in participants with CUD across different decision-making tasks (16). Thus, it is evident that activity of striatal and prefrontal cortical (PFC) regions is altered in CUD.

Functional MRI studies in individuals with GD have also found differential recruitment of PFC areas during reward-based tasks (3). The ventromedial PFC (vmPFC), an area activated during monetary reward tasks in healthy individuals that is important for reward processing, shows lower task-related activation in GD (17). On the Iowa Gambling Task, greater activity in individuals with GD during high-risk choices has been reported in the right caudate, OFC,

vmPFC, superior frontal gyrus (SFG), amygdala, and hippocampus (18). Furthermore, lower activity in the right ventrolateral PFC (vlPFC) has been linked to higher levels of perseveration on a PRL task (19). These findings point to altered reward processing in GD and suggest the involvement of cortical areas such as the vmPFC and OFC as well as subcortical structures; several areas overlap with those also affected in CUD. However, the neural substrates underlying RL in GD and CUD are not clearly defined. In rats, stickiness positively correlated with resting-state fMRI activity between the medial OFC (mOFC), PFC and subcortical structures (20,21). In humans, the link between RL behavior and neural activity in these clinical populations has not yet been established.

Here, we present a re-analysis of a previously published dataset (22) using novel computational methods. Individuals with CUD, GD and controls completed a PRL task in an fMRI scanner. In the previous publication arising from this dataset, conventional PRL measures were calculated and compared between the groups. There, it was reported that a behavioral variable reflecting the perseveration error rate was increased in CUD, with no differences observed in the GD group. Additionally, both patient groups had lower vlPFC activation when shifting responding following a reversal. In the new analysis presented here, RL models are employed to reveal latent processes underlying behavior on the PRL task, via a potentially more sensitive trial-by-trial approach. Through the fMRI data, the RL parameters can be linked to their associated neural substrates. To our knowledge, no previous studies have analyzed PRL data from GD patients using RL models. Based on our recent work that showed the concept of stickiness was critical for dissociating other disorders of compulsivity (13), we hypothesized that individuals with GD and CUD would show impairments in stickiness, and that stickiness would be greater in CUD. Neurally, we predicted that activity in the OFC would be linked to

the reward learning rate, and that medial PFC and dorsal striatal activity would reflect the stickiness parameter.

Methods

Participants

Fifty-six participants took part in this study. These comprised: 18 healthy control subjects who did not meet any of the criteria for an Axis I or II disorder; 18 subjects who met the DSM-IV-TR criteria for Pathological Gambling and 20 individuals that met the criteria for Cocaine Dependence. Here, we use the terms Cocaine Use Disorder and Gambling Disorder, which are the current nomenclature in the DSM-5, rather than Cocaine Dependence and Pathological Gambling, respectively, as used in the DSM-IV-TR (1). Basic behavioral data in association with fMRI findings from this study have previously been published (22). The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by the Ethics Committee for Research in Humans, University of Granada, Spain, approval number CEIH 2009/052. Participants signed an informed consent form to confirm their voluntary participation and were all equally reimbursed for their participation. Written informed consent was obtained from all participants. Please see **Supplementary Materials** for further information on participant recruitment.

Probabilistic Reversal Learning Task

This task was similar to the PRL task used by (15). Two abstract, colored stimuli were presented on the right and left side of the visual display. Stimulus location was randomized. At the beginning of the tasks, everyone was informed that one stimulus was the 'correct' stimulus

(CS+), and the other stimulus was the 'incorrect' stimulus (CS−). Subjects had to learn the correct and incorrect stimulus through a trial-and-error approach. The CS+ resulted in a reward on only 85% of the trials, whereas the CS− was rewarded 15% of the time. Following 10 to 15 correct responses, the contingencies were reversed. All participants were trained on the PRL task outside the scanner before the initial scan, for which different stimuli were used. During scanning, there were three consecutive blocks that consisted of 10 discriminations (9 reversals), with a duration of 11 min per block.

Magnetic-resonance-compatible liquid-crystal display goggles were used to present the stimuli (Resonance Technology Inc., Northridge, CA, USA). All responses were recorded using the Evoke Response Pad System (Resonance Technology Inc.). This button box was located on the subject's chest. The duration of stimulus presentation was 2000 ms. If participants failed to respond during this time, a 'too late' message was presented. Following a 'correct' response, a green smiley face was presented, and following an 'incorrect' response, a red sad face was shown. Feedback was presented for 500 ms, during which time the stimulus remained on the screen. Following feedback presentation, there was a variable inter-trial interval, which was adjusted by the program, for a final interstimulus interval duration between stimuli of 3253 ms. This interstimulus interval duration was selected to enable a precise desynchronization from the repetition time (2000 ms).

Reinforcement learning modeling

The PRL data was modelled with RL models using a hierarchical Bayesian approach. Six different models were run to test different combinations of model parameters, implemented through Stan (version 2.26.1) (23).

Q values were updated on a trial-by-trial basis according to the following equation:

$$Q_{t+1}(c_t) = Q_t(c_t) + \alpha \times \left(r_t - Q_t(c_t)\right) \quad (1)$$

$Q_{t+1}(c_t)$ is the expected value for the next trial based on the stimulus that is chosen on the current trial, $Q_t(c_t)$ is the expected value of the choice taken on the current trial, α is the learning rate and r_t is the reinforcement on trial t (1 for reward and 0 for punishment). The learning rate influences how much the subject updates the Q value based on the prediction error $r_t - Q_t(c_t)$, with higher α driving faster learning.

The probability of making one of two choices given the Q values for each was calculated using the softmax decision rule:

$$P\left(c_t = L \mid Q_t(L), Q_t(R)\right) = \frac{e^{Q_t(\mathrm{L})\beta}}{e^{Q_t(\mathrm{L})\beta} + e^{Q_t(\mathrm{R})\beta}} \quad (2)$$

$Q_t(L)$ and $Q_t(R)$ are the Q values of the left and right stimuli, and β is the reinforcement sensitivity parameter, which determines to what extent the subject is driven by its reinforcement history (versus random choice). Six models were tested and the parameters from the winning model were subsequently used for data simulation. Further information on these methods can be found in the **Supplementary Materials**.

First-level models

Information on image acquisition and pre-processing can be found in the **Supplementary Materials.** First-level linear models were fit through FEAT (FSL) (24). A first-level model was fit for each run and included the following event types: (1) reward Expected Value (EV), (2) positive Reward Prediction Error (RPE), (3) negative RPE, (4) punishment EV, (5) positive Punishment Prediction Error (PPE), (6) negative PPE and (7) response/feedback presentation. The RPE is when the prediction error is greater than 0 (when the outcome is better than the expected value) and is positive when there is a reward, and negative if the reward is omitted.

The PPE is the prediction error is below 0 (outcome is worse than the expected value). Similarly to the RPE, it is positive if there is a reward, and negative when these is no reward. RPEs take values between 0 and 1, whereas PPEs are between 0 and -1. EVs and prediction errors (PEs) were extracted for each trial from the winning Q-learning model. Explanatory variables 1–6 were based on the extracted values of prediction error and expected Q values as calculated in equation 1. The model was based on an analysis presented previously (25). Event types 1 and 4 were fitted during stimulus presentation, whereas 2, 3, 4 and 6 were fitted during feedback presentation. These first-level model regressors represent trial-level measures, whereas the RL parameters introduced in the previous section are subject-level measures. EVs, RPEs and PPEs were added as parametric modulators for the respective event types. Six movement parameters (x, y, z, pitch, roll, yaw) were incorporated into the model, which resulted from the image realignment to control for movement artefacts.

Higher-level models

The first-level models were averaged across the three runs for each subject, resulting in the second-level models. Third-level mixed-effects whole-brain analyses involving one-factor three-level ANOVAs with *post hoc* t-tests and cluster thresholding with a Z threshold of ± 3.1 and $p<0.05$ were used to investigate the contrasts for each event type (26). Subsequently, an analysis of covariance (ANCOVA) was run as an additional exploratory analysis. In the ANCOVA, model parameters from the best-fitting RL model were extracted for each subject and included as predictors. The aim of this analysis was to investigate group differences in the correlation between activity in a given region and a RL parameter (i.e., a group×RL parameter interaction). RL parameters were also correlated with BOLD signal from all participants, regardless of group. FSLeyes was used to generate figures (27). In all figures, the right and left

sides are inverted from the observer's perspective (according to standard radiological convention).

Results

Demographic information

There were no significant differences in age, gender, IQ, handedness, or years of education between the groups (**Table 1**) (Verdejo-Garcia et al. 2015).

Selecting the winning model

Table 2 reports the results from the six RL models tested and model comparison measures. Satisfactory model convergence was confirmed, as all parameters and contrasts had a potential scale reduction factor of less than 1.1, with the maximum value being 1.006.

The winning model (model 6) contained five parameters: the reward learning rate α_{rew}, representative of how quickly an individual updates (increases) Q values in response to positive feedback; the punishment learning rate α_{pun}, reflecting how quickly an individual updates (decreases) the Q value following punishment; reinforcement sensitivity β, also known as the exploitation vs exploration or inverse temperature parameter; stimulus stickiness κ_{stim}, which is the tendency to select the same stimulus regardless of outcome, and side stickiness κ_{side}, the tendency to select the same side regardless of outcome.

Reinforcement learning results

Figure 1 shows results of the hierarchical Bayesian RL analysis. Neither the reward learning rate nor the punishment learning rate were affected in GD or CUD when compared with healthy controls. However, there was evidence that the reward learning rate α_{rew} was lower in the CUD group than the GD group (difference in parameter per-group mean, posterior 75% HDI

excluding zero). Reinforcement sensitivity was lower in the CUD group compared to the GD group, reflecting more exploratory behavior in CUD, as well as higher κ_{stim} in the CUD group compared to the GD group (group differences, $0 \notin$ 75% HDI). Side stickiness, meanwhile, was not different in either patient group compared to the control group (no group differences, $0 \in$ 75% HDI). There was evidence for lower stimulus stickiness at 75% HDI in the GD group compared to HCs (group difference, $0 \notin$ 75% HDI). There were no differences in the CUD group when compared to the control group (no group differences, $0 \in$ 75% HDI). To summarize, we found evidence for the stimulus stickiness parameter κ_{stim} being lower in the GD group compared to HCs. No differences at 95% HDI were observed. We note that 95% HDI provides stronger evidence for there being group differences than 75% HDI, however, 75% HDI is also considered to provide sufficient evidence and has been used as a threshold in previous studies (28).

Simulations

The parameters from the winning RL model were used to simulate the behavioral data and determine whether this model could replicate the behavior observed initially via raw data measures. When these data were analyzed using a conventional approach to extract raw data measures such as the proportion of correct responses trials to criterion and number of perseverative responses, no statistically significant differences between the groups were found. These findings thus align with the results for the conventional behavioral measures presented in (22), suggesting that the model was able to reproduce the behavioral dynamics on this task. The results can also be seen in the **Supplementary Materials.**

Brain activity during reward and punishment expected value tracking in Gambling Disorder

The model fitted to the task-based fMRI data included seven explanatory variables, as above: (1) reward EV; (2) positive RPE; (3) negative RPE; (4) punishment EV; (5) positive PPE; (6) negative PPE and (7) response/feedback presentation. We found differences in the neural responses to reward and punishment expected value in the GD group compared to controls. Specifically, we observed that when tracking reward EV, that individuals with GD had greater activations in the amygdala, hippocampus, parahippocampal gyrus, lateral occipital cortex, superior, inferior, and middle temporal gyri, posterior cingulate gyrus as well as the precuneus than HCs (**Figure 2**, **Table 3**). These effects were only observed in the left hemisphere.

For punishment EV, we observed the opposite trend: individuals with GD showed lower activity in the superior parietal lobule, pre- and postcentral gyri, precuneus, parietal operculum, supramarginal gyrus and angular gyrus compared to control subjects (**Figure 3**, **Table 4**). Activations were seen in both hemispheres but were more pronounced in the right hemisphere.

Neural signal to positive and negative punishment prediction errors is altered in Cocaine Use Disorder

We observed aberrant neural responses in CUD as well, specifically in response to positive and negative PPEs. Compared to control participants, individuals with CUD exhibited lower activity in the paracingulate gyrus and left SFG in response to positive PPEs. Conversely, individuals with CUD showed greater activity in the left SFG and MFG in response to negative PPEs (**Figures 4, 5**; **Tables 5, 6**, respectively).

Further results on the neural responses to feedback presentation can be found in the **Supplementary Materials.**

Whole-brain correlation analyses

The five parameters from the winning RL model were used in a whole-brain correlation analysis to identify whether they correlated with the BOLD signal during each event type in any of the brain regions. This was done to identify the brain regions underlying RL parameters. The first analysis related the parameters to activity from all subjects.

This analysis highlighted that the α_{rew} parameter correlated negatively with activity in the cingulate and paracingulate gyri, IFG, middle and superior temporal gyri, insular cortex, and mOFC during reward EV tracking as well as responses to positive PPEs. This parameter also correlated negatively with activity in the putamen, mOFC, and insula during positive RPEs.

Next, an ANCOVA was run to compare task-based activity among the different groups. In the GD group, α_{rew} correlated more strongly with activity in the SFG, MFG, postcentral gyrus during reward EV tracking compared to the other two groups (**Figure S3**). In the CUD group, the correlation between α_{rew} and activity during the positive PPE was greater in the frontal pole, SFG, cingulate and paracingulate gyri compared to the HC and GD groups (**Figure S4**).

In both patient groups, stimulus stickiness (κ_{stim}) had a stronger positive correlation with activity in the right MFG and IFG during response/feedback presentation compared to control participants, suggesting that there is greater activity in these areas in patients when repeating a response regardless of previous outcomes (**Figure 6**). No other correlations with RL parameters were found.

Discussion

In this study, we examined RL processes during a classic test of behavioral flexibility (PRL) in individuals with GD and CUD. Our computational modeling approach enabled the assessment of how both value-based (learning rates, reinforcement sensitivity) and value-free (stimulus and side stickiness) contributed to choice behavior. The key behavioral result was that individuals with GD showed reduced choice repetition (stimulus stickiness), irrespective of the feedback received, suggestive of a maladaptive exploratory pattern. Reduced stimulus stickiness in GD contrasts with our recent observation of greater choice repetition in SUD, regardless of reinforcement (13). Our findings also extend the results presented in (22), which found a higher perseveration error rate in CUD and no differences in GD. We thus demonstrate that the use of RL modelling can provide a novel insight into PRL data and may help to explain which parameters contribute to differences in conventional measures. Stimulus stickiness (a form of choice repetition) may therefore present a novel way of dissociating compulsive disorders, in this case GD and CUD. However, we note that group differences were only observed at 75% HDI, but not at 95%. Furthermore, the sample sizes were relatively small.

We provide a novel and unexpected insight into how RL parameters are affected in GD – that stimulus stickiness was reduced in this group. A similar reduction in stimulus stickiness has also been observed in another compulsive disorder, OCD (13). However, in GD, the reduction in stimulus stickiness was accompanied by slightly higher levels of side stickiness κ_{side} (below 75% HDI), whereas in OCD there was additionally a mild reduction in side stickiness (13). In other words, the computational profile of GD and OCD appears to be distinct. Perseveration is not a unitary construct: side stickiness may be representative of motor perseveration, whereas stimulus stickiness reflects stimulus perseveration. Side stickiness may therefore represent excessive motor perseveration. In contrast, lower stimulus stickiness may reflect another form of behavioral inflexibility that is overly exploratory yet outcome insensitive. Low stimulus

stickiness in GD detected during trial-and-error learning in a laboratory setting may therefore reflect a real-life increase in exploration of choices in an attempt to identify an optimal strategy, e.g., tracking new stimuli in a casino game. Overall, value-free contributors to choice behavior have allowed for novel dissociations of GD, OCD, and SUD, and point to a possible computational fingerprinting, which could eventually be useful for informing psychiatric classification.

At the neural level, group differences were also observed during ongoing RL processes. Differences in brain activity when tracking reward and punishment EVs were seen in participants with GD. In these individuals, there was greater activity in response to reward EVs in areas including the amygdala, hippocampus and cingulate gyrus compared to HCs. When tracking punishment EV, on the other hand, there was lower activity in regions such as the postcentral gyrus, superior parietal lobule and occipital areas, suggesting that individuals with GD differentially track EVs of stimuli in their surroundings in favor of reward-related expectancies. In the CUD group, there also appeared to be an altered balance in RL, instead with lower responses to positive PPEs and greater responses to negative PPEs in the SFG and neighboring regions compared to control participants, which suggests preferential processing of punishment. This aligns with our recent finding that individuals with SUD show greater punishment learning rates (13). We highlight the application of Bayesian statistics to the RL modelling data compared to ANOVAs for the imaging data as a limitation of our analysis. In summary, there appear to be uniquely aberrant neural signals in each patient group when tracking value-related information important for RL processes.

By linking the computational modeling parameters to the fMRI data, we also identified regions involved in the modulation of RL measures, which has not been investigated in previous human

studies. We found that the learning rate parameter for reward (α_{rew}) was correlated with areas that responded to RPEs and PPEs, including the SFG, MFG, cingulate and paracingulate gyri. Therefore, these regions appear to be of key importance for RL and are likely to be involved in the modulation of the reward learning rate (α_{rew}). The SFG and ACC are key areas that have been shown to be involved in error and action monitoring, providing support for their involvement in reward learning (29). Moreover, a meta-analysis including 35 studies reported that these areas are consistently activated when there is a prediction error (30).

At least two previous studies have reported reduced learning rates, reinforcement sensitivity and greater stimulus stickiness in individuals with SUD compared with HCs (13,31). In the present study, meanwhile, we observed lower reward learning rates and higher stimulus stickiness in CUD only when contrasted with GD. Duration of substance abuse may be a key factor underlying the less pronounced RL results in CUD when compared to these two previous studies. Whereas the CUD sample in the present study had an average duration of substance use of 3.7 years (22), the participants with SUD in previous studies reporting more pronounced RL deficits had been using for an average of 11.7 (6) and 13.7 years (31). Additionally, a criterion in our study was abstinence, which was not the case in the other two investigations. These differences in sample suggest longer exposure to substances may have more pronounced effects on RL processes, possibly due to neurotoxicity, and may therefore help reconcile the RL findings between these studies. As GD itself does not involve substance use, we would not expect the same magnitude or mechanism of change in RL effects related to disease duration (which was 2.2 years for GD in our sample). At the same time, such contrasts between GD and SUD may inform which aspects of RL in SUD are more or less likely to be tied to neurotoxic effects.

Based on the neural results presented here, individuals with GD appear to be less sensitive to punishment EV but more sensitive to reward EV than controls. A study of performance on a two-choice lottery task found that choice behavior in GD patients was less sensitive to EVs for both reward and punishment, with this group using information about magnitude and probability information less than HCs (32). Thus, attenuated responses to punishment appear to be common across tasks in GD. Although greater sensitivity to reward was observed in our study compared to lower levels reported in Limbrick-Oldfield et al. (2021), this may have been because different behavioral paradigms were used. Consistent with our findings, a previous study employing a card-guessing task, participants with GD had greater neural responses in the VS and OFC when tracking reward EV (33). Overall, these studies suggest that GD patients show altered responses to reinforcement tracking and are less sensitive to punishment.

In individuals with SUD, reduced responses to PEs in the VS and mOFC on the IGT have been reported previously (34). In a separate study using electroencephalography, impaired RPE signaling in CUD was also found (35). In contrast, we found greater responses to PPEs, rather than reduction in RPEs. Following cocaine abstinence in individuals with CUD, enhanced signals to positive PEs, regardless of whether reward or punishment was predicted, have been observed (36). Although we report reduced activity following positive PPEs, this may be because we separated reward and punishment PEs and suggests that the two PEs are differentially altered in CUD. Altered responses to PE related to both reward and punishment could be a contributor to compulsive drug use, as it persists despite negative outcomes. In patients with OCD, RPE responses were altered in the nucleus accumbens and anterior cingulate cortex, further highlighting that RL can be used to distinguish disorders of compulsivity, both through behavior and its associated neural substrates (25).

We report that stimulus stickiness (κ_{stim}) was positively correlated with activity in the dorsolateral PFC (dlPFC) and ventrolateral PFC (vlPFC), areas important for cognitive control, including conflict monitoring and motor inhibition, respectively (37). In the results presented here, patients with GD and CUD showed a stronger positive correlation with stimulus stickiness (κ_{stim}) in these regions. This result was contrary to our expectations and previous studies, as it was predicted that stickiness would be related to *reduced* activity in these regions. A possible interpretation of this finding is that stimulus stickiness reflects bias towards one of the presented stimuli, ideally the majority reinforced one, and that the MFG and IFG are active in order to overcome this response following a reversal. However, this hypothesis would need to be explored further in future studies.

It has been demonstrated previously that both the dlPFC and vlPFC are affected in GD and CUD (38,39); here we provide a novel computational mechanism pertinent to compulsions that is linked to these regions in GD and CUD. Previous studies have demonstrated that response shifting on the PRL task is associated with vlPFC activation in control participants (15). Consistent with the present results, a prior analysis of this dataset showed the vlPFC was engaged during response shifting, yet both clinical groups showed lower vlPFC activity than HCs (22). Reduced vlPFC activity during shifting has been also reported in OCD patients (40). These findings from previous studies, however, focus on response shifting on certain trials, whereas our analysis investigated stickiness across all trials, reflecting an overall tendency. Additionally, stickiness represents repeated responses, rather than response shifts. In rats, it has been shown that side stickiness (stimulus stickiness was not studied) is correlated with activity in medial PFC and dorsal striatal regions (20). It is therefore possible that side and stimulus stickiness recruit different neural circuits, but this requires further analysis in the same species.

In summary, we provide novel behavioral and neural insights into GD through computational modeling of RL processes. Critically, we demonstrate that individuals with GD and CUD display perseverative behavior during PRL that differs both qualitatively and quantitatively, advancing the notion that compulsivity is not a unitary construct. We also provide evidence that individuals with GD and CUD display aberrant and opposing neural responses to rewards and punishments, in relation to expected value and PPEs. Furthermore, we link RL parameters to regions that may be involved in their modulation, which has not previously been investigated in the human literature. We demonstrate that RL modeling combined with fMRI may provide new insights into the mechanisms underlying compulsive disorders and therefore refine our understanding of compulsivity transdiagnostically.

Funding

The collection of data for this study was originally funded by the Spanish Ministry of Health / Plan Nacional Sobre Drogas and the Centre for Gambling Research at the University of British Columbia (UBC). R.N.C.'s research is supported by the UK Medical Research Council (MRC) (MR/W014386/1). K.Z. was supported by the Institute for Neuroscience at the University of Cambridge, the Alan Turing Institute, London and the Angharad Dodds John Bursary in Mental Health and Neuropsychiatry, Downing College, Cambridge. J.W.D. has received funding from GlaxoSmithKline and Boehringer Ingelheim Pharma GmbH and is a co-investigator on an MRC program grant (MR/N02530X/1). T.W.R. is also a co-investigator of the latter grant. J.W.K. was supported by an Angharad Dodds John Bursary in Mental Health and Neuropsychiatry. A.V.G. is supported by a Leadership Investigator Grant from the Australian National Health and Medical Research Council (GNT2009464). J.V-R., N.A-U. and C.S-M. have no financial or conflicts of interest to declare.

Declaration of Interests

The Centre for Gambling Research at UBC (L.C.) is funded by support from the Province of British Columbia and the British Columbia Lottery Corporation. J.W.D. has received research grants from Boehringer Ingelheim Pharma GmbH and GlaxoSmithKline and receives royalties from Springer Verlag. T.W.R. discloses consultancy with Cambridge Cognition; he receives editorial honoraria from Springer-Nature and Elsevier and a research grant from Shionogi. R.N.C. consults for Campden Instruments and receives royalties from Cambridge Enterprise, Routledge, and Cambridge University Press. K.Z., J.W.K. and A.V.G. have no conflicts of interest to declare.

Author contributions

KZ: conceptualization, software, methodology, validation, formal analysis, visualization, writing – original draft, writing – review & editing; JVR – conceptualization, methodology, investigation, data curation, writing – review & editing; LC – conceptualization, methodology, investigation, project administration, funding acquisition; NAU – conceptualization, methodology, investigation, data curation; CSM – conceptualization, software, methodology, investigation, formal analysis, writing – review & editing; RNC – conceptualization, software, supervision, formal analysis, validation, writing – review & editing; TWR – supervision, writing – original draft, writing – review & editing; JWD – conceptualization, writing – original draft, writing – review & editing, supervision; AVG – conceptualization, methodology, investigation, resources, writing – original draft, writing – review & editing, supervision, funding acquisition; JWK – conceptualization, methodology, writing – original draft, writing – review & editing, supervision.

Bibliography

1. APA. American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders (5th ed.). American Journal of Psychiatry. 2013.
2. Dalley JW, Everitt BJ, Robbins TW. Impulsivity, Compulsivity, and Top-Down Cognitive Control. Neuron. 2011 Feb 24;69(4):680–94.
3. Leeman RF, Potenza MN. Similarities and differences between pathological gambling and substance use disorders: A focus on impulsivity and compulsivity. Psychopharmacology (Berl) [Internet]. 2012 Jan 5 [cited 2022 Mar 29];219(2):469–90. Available from: https://link.springer.com/article/10.1007/s00213-011-2550-7
4. Jara-Rizzo MF, Navas JF, Rodas JA, Perales JC. Decision-making inflexibility in a reversal learning task is associated with severity of problem gambling symptoms but not with a diagnosis of substance use disorder. BMC Psychol [Internet]. 2020 Nov 10 [cited 2022 Mar 29];8(1):120. Available from: https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-020-00482-6
5. Perandrés-Gómez A, Navas JF, van Timmeren T, Perales JC. Decision-making (in)flexibility in gambling disorder. Addict Behav. 2021 Jan 1;112:106534.
6. Ersche KD, Roiser JP, Abbott S, Craig KJ, Mller U, Suckling J, et al. Response Perseveration in Stimulant Dependence Is Associated with Striatal Dysfunction and Can Be Ameliorated by a D2/3 Receptor Agonist. Biol Psychiatry. 2011 Oct 15;70(8):754–62.
7. Jentsch JD, Olausson P, De La Garza R, Taylor JR. Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. Neuropsychopharmacology [Internet]. 2002 Aug 2 [cited 2021 May 21];26(2):183–90. Available from: www.acnp.org/citations/
8. Schoenbaum G, Saddoris MR, Ramus SJ, Shaham Y, Setlow B. Cocaine-experienced rats exhibit learning deficits in a task sensitive to orbitofrontal cortex lesions. Eur J

Neurosci [Internet]. 2004 Apr [cited 2022 Dec 19];19(7):1997–2002. Available from: https://pubmed.ncbi.nlm.nih.gov/15078575/

9. Leppink EW, Redden SA, Chamberlain SR, Grant JE. Cognitive flexibility correlates with gambling severity in young adults. J Psychiatr Res. 2016 Oct 1;81:9–15.

10. 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 Jan 1;84:204–17.

11. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. IEEE Trans Neural Networks. 1998 Sep;9(5):1054–1054.

12. Zhukovsky P, Puaud M, Jupp B, Sala-Bayo J, Alsiö J, Xia J, et al. Withdrawal from escalated cocaine self-administration impairs reversal learning by disrupting the effects of negative feedback on reward exploitation: a behavioral and computational analysis. Neuropsychopharmacology [Internet]. 2019 Dec 1 [cited 2020 Jun 23];44(13):2163–73. Available from: https://www.nature.com/articles/s41386-019-0381-0

13. 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 (Berl) [Internet]. 2019 Aug 1 [cited 2021 Mar 16];236(8):2337–58. Available from: https://doi.org/10.1007/s00213-019-05325-w

14. Goldstein RZ, Alia-Klein N, Tomasi D, Zhang L, Cottone LA, Maloney T, et al. Is decreased prefrontal cortical sensitivity to monetary reward associated with impaired motivation and self-control in cocaine addiction? Am J Psychiatry. 2007;164(1):43–51.

15. Cools R, Clark L, Owen AM, Robbins TW. Defining the Neural Mechanisms of

Probabilistic Reversal Learning Using Event-Related Functional Magnetic Resonance Imaging. J Neurosci [Internet]. 2002 Jun 1 [cited 2021 Nov 22];22(11):4563–7. Available from: https://www.jneurosci.org/content/22/11/4563

16. Dom G, Sabbe B, Hulstijn W, Van Den Brink W. Substance use disorders and the orbitofrontal cortex: Systematic review of behavioural decision-making and neuroimaging studies. Br J Psychiatry [Internet]. 2005 Sep [cited 2022 Mar 31];187(3):209–20. Available from: https://www.cambridge.org/core/journals/the-british-journal-of-psychiatry/article/substance-use-disorders-and-the-orbitofrontal-cortex/5F1E8C169A1E665DC9447AA54C9BD945

17. Habib R, Dixon MR. Neurobehavioral evidence for the "Near-Miss" effect in pathological gamblers. J Exp Anal Behav [Internet]. 2010 May [cited 2022 Mar 31];93(3):313–28. Available from: https://pubmed.ncbi.nlm.nih.gov/21119848/

18. Power Y, Goodyear B, Crockford D. Neural Correlates of Pathological Gamblers Preference for Immediate Rewards During the Iowa Gambling Task: An fMRI Study. J Gambl Stud [Internet]. 2012 Dec 25 [cited 2022 Mar 31];28(4):623–36. Available from: https://link.springer.com/article/10.1007/s10899-011-9278-5

19. De Ruiter MB, Veltman DJ, Goudriaan AE, Oosterlaan J, Sjoerds Z, Van Den Brink W. Response Perseveration and Ventral Prefrontal Sensitivity to Reward and Punishment in Male Problem Gamblers and Smokers. Neuropsychopharmacol 2009 344 [Internet]. 2008 Oct 1 [cited 2022 Mar 31];34(4):1027–38. Available from: https://www.nature.com/articles/npp2008175

20. Zühlsdorff K, López-Cruz L, Dutcher EG, Jones JA, Pama C, Sawiak S, et al. Sex-dependent effects of early life stress on reinforcement learning and limbic cortico-striatal functional connectivity. Neurobiol Stress [Internet]. 2023 Jan 1 [cited 2022 Dec 19];22:100507. Available from:

https://linkinghub.elsevier.com/retrieve/pii/S2352289522000820

21. Zuhlsdorff K. Investigating reinforcement learning processes in depression and substance use disorder: translational, computational and neuroimaging approaches. 2022 Jul 19 [cited 2022 Dec 19]; Available from: https://www.repository.cam.ac.uk/handle/1810/343811

22. 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 [Internet]. 2015 Aug 1 [cited 2020 Dec 8];207(2):158–64. Available from: https://pubmed.ncbi.nlm.nih.gov/26045346/

23. Stan Development Team. Stan Modeling Language User's Guide and Reference Manual, Version 2.19.2. Interact Flow Model Lang. 2020;

24. Woolrich MW, Ripley BD, Brady M, Smith SM. Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. Neuroimage. 2001 Dec 1;14(6):1370–86.

25. 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. Psychopharmacol 2019 2368 [Internet]. 2019 Jun 14 [cited 2021 Oct 11];236(8):2325–36. Available from: https://link.springer.com/article/10.1007/s00213-019-05292-2

26. Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM. Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage. 2004 Apr 1;21(4):1732–47.

27. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. In: NeuroImage. 2004.

28. Rygula R, Clarke HF, Cardinal RN, Cockcroft GJ, Xia J, Dalley JW, et al. Role of

central serotonin in anticipation of rewarding and punishing outcomes: Effects of selective amygdala or orbitofrontal 5-HT Depletion. Cereb Cortex [Internet]. 2015 Sep 1 [cited 2021 Mar 17];25(9):3064–76. Available from: https://pubmed.ncbi.nlm.nih.gov/24879752/

29. Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD. Conflict monitoring versus selection for-action in anterior cingulate cortex. Nature [Internet]. 1999 Nov 11 [cited 2021 Jan 14];402(6758):179–81. Available from: www.nature.com

30. Garrison J, Erdeniz B, Done J. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev. 2013 Aug 1;37(7):1297–310.

31. Lim TV, Cardinal RN, Bullmore ET, Robbins TW, Ersche KD. Impaired learning from negative feedback in stimulant use disorder: Dopaminergic modulation. Int J Neuropsychopharmacol [Internet]. 2021 Jul 1 [cited 2021 Jul 7]; Available from: https://academic.oup.com/ijnp/advance-article/doi/10.1093/ijnp/pyab041/6312679

32. Limbrick-Oldfield EH, Cherkasova M V., Kennedy D, Goshko CB, Griffin D, Barton JJS, et al. Gambling disorder is associated with reduced sensitivity to expected value during risky choice. J Behav Addict [Internet]. 2021 Jan 1 [cited 2022 Dec 22];9(4):1044. Available from: /pmc/articles/PMC8969736/

33. Van Holst RJ, Veltman DJ, Bchel C, Van Den Brink W, Goudriaan AE. Distorted Expectancy Coding in Problem Gambling: Is the Addictive in the Anticipation? Biol Psychiatry. 2012 Apr 15;71(8):741–8.

34. Tanabe J, Reynolds J, Krmpotich T, Claus E, Thompson LL, Du YP, et al. Reduced Neural Tracking of Prediction Error in Substance-Dependent Individuals. Am J Psychiatry [Internet]. 2013 Nov 1 [cited 2022 Dec 22];170(11):1356. Available from: /pmc/articles/PMC4426095/

35. Parvaz MA, Konova AB, Proudfit GH, Dunning JP, Malaker P, Moeller SJ, et al. Impaired Neural Response to Negative Prediction Errors in Cocaine Addiction. J Neurosci [Internet]. 2015 [cited 2022 May 11];35(5):1872. Available from: /pmc/articles/PMC4315825/

36. Wang JM, Zhu L, Brown VM, De La Garza R, Newton T, King-Casas B, et al. In Cocaine Dependence, Neural Prediction Errors During Loss Avoidance Are Increased With Cocaine Deprivation and Predict Drug Use. Biol Psychiatry Cogn Neurosci Neuroimaging. 2019 Mar 1;4(3):291–9.

37. Levy BJ, Wagner AD. Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. Ann N Y Acad Sci [Internet]. 2011 [cited 2023 Feb 17];1224(1):40. Available from: /pmc/articles/PMC3079823/

38. Goldstein RZ, Volkow ND. Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. Vol. 12, Nature Reviews Neuroscience. 2011.

39. Raimo S, Cropano M, Trojano L, Santangelo G. The neural basis of gambling disorder: An activation likelihood estimation meta-analysis. Vol. 120, Neuroscience and Biobehavioral Reviews. 2021.

40. Remijnse PL, Nielen MMA, Van Balkom AJLM, Cath DC, Van Oppen P, Uylings HBM, et al. Reduced Orbitofrontal-Striatal Activity on a Reversal Learning Task in Obsessive-Compulsive Disorder. Arch Gen Psychiatry [Internet]. 2006 Nov 1 [cited 2022 Apr 29];63(11):1225–36. Available from: https://jamanetwork.com/journals/jamapsychiatry/fullarticle/668224

Table 1. Demographic information.

	Healthy Controls (n=18)	**Gambling Disorder (n=18)**	**Cocaine Use Disorder (n=20)**	**Group Comparisons**
Mean age (SD)	31.2 (4.7)	33.6 (8.0)	34.3 (6.9)	$F(2,54)=1.43$, p=0.35
Gender (Female)	1	2	1	$X^2(2,56)=0.59$, p=0.75
Verbal IQ (SD)	106.9 (9.0)	102.7 (7.4)	100.9 (7.6)	$F(2,54)=2.31$, p=0.082
Years of Education (SD)	10.6 (1.9)	10.3 (2.1)	9.8 (1.7)	$F(2,54)=1.37$, p=0.47
Handedness (L)	1	1	4	$X^2(2,56)=2.80$, p=0.25

L, left; SD, standard deviation.

Table 2. Model comparison summary. Models were assumed to be equiprobable *a priori*.

Model	**Rank**	**Parameters**	**Log marginal likelihood**	**Log posterior P**
1	4	α, β	-13159.06	-573.18
2	5	α, β, κ_{stim}	-13018.42	-432.54
3	3	α_{rew}, $\alpha_{non\text{-}rew}$, β	-13130.08	-544.21
4	2	α_{rew}, $\alpha_{non\text{-}rew}$, β, κ_{stim}	-13003.72	-417.85
5	6	α_{rew}, $\alpha_{non\text{-}rew}$, β, κ_{side}	-13168.48	-582.60
6	1	α_{rew}, $\alpha_{non\text{-}rew}$, β, κ_{side}, κ_{stim}	-11139.35	0.000

Table 3. Summary of peak fMRI activity for the reward EV controls-vs-GD contrast.

Name	**BA**	**Side**	**MNI coordinates (X, Y, Z)**	**Number of voxels**	**Volume (mm^3)**	**Mean *Z* statistic**
Middle temporal gyrus	21	L	-57, -6, -17	365	5893	3.63
Precuneus	7	L	-3, -66, 33	224	3617	3.68
Cingulate gyrus	24, 32	L	-9, -50, 27	182	2939	3.71
Superior temporal gyrus	22, 42	L	-46, -14, -8	112	1808	3.49
Lateral occipital cortex	19	L	-57, -62, -6	97	1566	3.79
Hippocampus	28	L	-21, -10, -24	81	1308	3.45
Amygdala		L	-23, -5, -17	68	1098	3.41

Parahippocampal gyrus	27	L	-17, -10, -24	59	953	3.38
Inferior temporal gyrus	20	L	-57, -57, -13	29	468	3.42

Whole-brain analysis involving one-sample *t* tests with cluster thresholding with a *Z* threshold of 3.1 and p<0.05. The areas indicated show greater activity in participants with CUD than control participants. BA, Brodmann area; MNI, Montreal Neurological Institute template.

Table 4. Summary of peak fMRI activity for the punishment EV controls-vs-GD contrast.

Name	**BA**	**Side**	**MNI coordinates (X, Y, Z)**	**Number of voxels**	**Volume (mm^3)**	**Mean z-statistic**
Postcentral gyrus	1, 2, 3	R	56, -14, -33	1228	19827	2.99
Postcentral gyrus	1, 2, 3	L	-62, -21, -33	408	6588	3.02
Precentral gyrus	4	R	43, -14, 45	865	13966	2.95
Precuneus	7	R	10, -51, 56	555	8961	2.90
Precuneus	7	L	-6, -48, 56	403	6507	2.90
Superior parietal lobule	7	R	28, -44, 59	524	8461	3.04
Supramarginal gyrus	40	R	56, -18, 31	254	4101	2.88
Supramarginal gyrus	40	L	-63, -26, 31	258	4166	3.02
Lateral occipital cortex	19	R	17, -79, 41	242	3907	2.88
Lateral occipital cortex	19	L	-14, -83, 41	181	2922	2.75
Parietal operculum cortex	40, 43	R	1.5, -33, 21	106	1711	2.83
Parietal operculum cortex	40, 43	L	-51, -33, 21	191	3084	2.98

Whole-brain analysis involving one-sample *t* tests with cluster thresholding with a *Z* threshold of 3.1 and p<0.05. The areas indicated show lower activity in participants with CUD than control participants. BA, Brodmann area; MNI, Montreal Neurological Institute template.

Table 5. Summary of peak fMRI activity for the positive PPE controls-vs-CUD contrast.

Name	**BA**	**Side**	**MNI coordinates (X, Y, Z)**	**Number of voxels**	**Volume (mm^3)**	**Mean z-statistic**
Superior frontal gyrus	8, 9	L	-10, 13, 53	149	2406	2.81
Paracingulate gyrus	32	L	-8, 20, 43	132	2131	2.80
Paracingulate gyrus	32	R	4, 11, 47	36	581	2.71

Whole-brain analysis involving one-sample *t* tests with cluster thresholding with a *Z* threshold

of 3.1 and p<0.05. The areas indicated show lower activity in participants with CUD than control participants. BA, Brodmann area; MNI, Montreal Neurological Institute template.

Table 6. Summary of peak fMRI activity for the punishment PPE controls-vs-CUD contrast.

Name	**BA**	**Side**	**MNI coordinates (X, Y, Z)**	**Number of voxels**	**Volume (mm^3)**	**Mean z-statistic**
Superior frontal gyrus	8, 9	L	-57, -6, -17	71	1146	3.41
Middle frontal gyrus	8, 9	L	-3, -66, 33	70	1130	3.41

Whole-brain analysis involving one-sample *t* tests with cluster thresholding with a *Z* threshold of 3.1 and p<0.05. The areas indicated show greater activity in participants with CUD than control participants. BA, Brodmann area; MNI, Montreal Neurological Institute template.

Figure 1. Results from the hierarchical Bayesian winning RL model, showing differences in group mean parameters. GD, Gambling Disorder; CUD, Cocaine Use Disorder; HC, healthy controls; Reinf. sens, reinforcement sensitivity; Stim, stimulus; HDI, highest posterior density interval. Orange indicates 0 ∉ 75% HDI.

Figure 2. Reward EV tracking: differences between healthy controls and participants with GD (MNI coordinates: Y=-18 to -11). Activity was higher in the GD group in the areas indicated. Color bar on the right-hand side represents the *t* statistic.

Figure 3. Punishment EV tracking: differences between healthy controls and participants with GD (MNI coordinates: Y=-24 to -17). Activity was lower in the GD group in the areas indicated. Color bar on the right-hand side represents *t*.

Figure 4. Response to positive PPEs: differences between healthy controls and participants with CUD (MNI coordinates: X=-5, Y=17, Z=48). Activity was lower in the CUD group in the areas indicated. Color bar on the right-hand side represents *t*.

Figure 5. Response to negative PPEs: differences between healthy controls and participants with CUD (MNI coordinates: X=-31, Y=30, Z=56). Activity was higher in the CUD group in the areas indicated. Color bar on the right-hand side represents *t*.

Figure 6. Top: Areas that have a stronger positive correlation with κ_{stim} in the GD group than in healthy controls (MNI coordinates: X=48, Y=29, Z=22). Bottom: Areas that have a stronger positive correlation with κ_{stim} in the CUD group than in healthy controls (MNI coordinates: X=48, Y=29, Z=20). Color bar on the right-hand side represents *t*.