Attenuated directed exploration during reinforcement learning in gambling disorder

Wiehler, A.[1,2], Chakroun, K.[1], Peters, J.[1,3]

[1]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

[2]Institut du Cerveau et de la Moelle épinière (ICM), INSERM U 1127, CNRS UMR 7225, Sorbonne Universités Paris, France.

[3]Department of Psychology, Biological Psychology, University of Cologne, Cologne, Germany.

Correspondence: antonius.wiehler@gmail.com, jan.peters@uni-koeln.de

Abstract

Gambling disorder is a behavioral addiction associated with impairments in decision-making and reduced behavioral flexibility. Decision-making in volatile environments requires a flexible trade-off between exploitation of options with high expected values and exploration of novel options to adapt to changing reward contingencies. This classical problem is known as the exploration-exploitation dilemma. We hypothesized gambling disorder to be associated with a specific reduction in directed (uncertainty-based) exploration compared to healthy controls, accompanied by changes in brain activity in a fronto-parietal exploration-related network.

Twenty-three frequent gamblers and nineteen matched controls performed a classical four-armed bandit task during functional magnetic resonance imaging. Computational modeling revealed that choice behavior in both groups contained signatures of directed exploration, random exploration and perseveration. Gamblers showed a specific reduction in directed exploration, while random exploration and perseveration were similar between groups.

Neuroimaging revealed no evidence for group differences in neural representations of expected value and reward prediction errors. Likewise, our hypothesis of attenuated fronto-parietal exploration effects in gambling disorder was not supported. However, during directed exploration, gamblers showed reduced parietal and substantia nigra / ventral tegmental area activity. Cross-validated classification analyses revealed that connectivity in an exploration-related network was predictive of clinical status, suggesting alterations in network dynamics in gambling disorder.

In sum, we show that reduced flexibility during reinforcement learning in volatile environments in gamblers is attributable to a reduction in directed exploration rather than an increase in perseveration. Neuroimaging findings suggest that patterns of network connectivity might be more diagnostic of gambling disorder than univariate value and prediction error effects. We provide a computational account of flexibility impairments in gamblers during reinforcement learning that might arise as a consequence of dopaminergic dysregulation in this disorder.

Keywords

Gambling disorder; reward; reinforcement learning; exploration-exploitation; perseveration; fMRI

Introduction

Gambling disorder (GD) is defined in the DSM-5 as "Persistent and recurrent problematic gambling behavior leading to clinically significant impairment or distress" (American Psychiatric Association, 2013). It has a worldwide lifetime prevalence of around 1% (Kessler et al., 2008; Lorains, Cowlishaw, & Thomas, 2011). In the DSM-5, it is the only behavioral addiction classified in the category of substance use and addictive disorders. This classification reflects the considerable overlap in the behavioral and neural correlates of gambling disorder with substance-based addictions, which is striking since no pharmacological agent is involved (Goudriaan, Brink, & Holst, 2019).

Decision-making impairments in gambling disorder include increased impulsivity in inter-temporal choice tasks (steeper temporal discounting, reflecting an increased preference for smaller-but-sooner over larger-but-later rewards) and, though somewhat less consistently, increased risk-taking (Wiehler & Peters, 2015). In line with these behavioral findings, activity in reward-related brain regions, including the ventral striatum and medial prefrontal cortex, has repeatedly been found to differ between healthy controls and participants with GD. However, the direction of these differences shows considerable inconsistencies between studies (Balodis et al., 2012; Clark, Boileau, & Zack, 2019; Leyton & Vezina, 2012), which has been suggested to be due to contextual (Miedl, Büchel, & Peters, 2014) and task-specific effects (Clark et al., 2019; Leyton & Vezina, 2012).

Gambling disorder is also associated with cognitive impairments predominantly reflected in reduced behavioral flexibility. For example, gamblers show impaired performance in the Stroop task and increased perseveration following rule changes in the Wisconsin Card Sorting Task (van Timmeren, Daams, van Holst, & Goudriaan, 2018). Similar impairments are observed in reversal learning tasks. Here, participants learn to select the stimulus with the higher reinforcement rate. Contingencies then reverse during the experiment, requiring participants to adapt to these changes, and gamblers make more perseveration errors following reversals (Boog et al., 2014; de Ruiter et al., 2009). More generally, such reward learning tasks can be understood as entailing a trade-off between exploration and exploitation. Agents repeatedly have to decide between exploiting choice options with known expected values and exploring other options with unknown, but potentially higher values. Reversal learning is a special case of this: After a reversal, the expected value of the formerly superior option is decreasing, requiring participants to explore the previously inferior option. Note that in reversal learning, exploration cannot readily be dissociated from perseveration behavior.

The exploration-exploitation trade-off has been extensively studied in psychology and cognitive neurosciences, ranging from foraging studies in animals to human psychology and computational modeling studies (Cohen, McClure, & Yu, 2007; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Mehlhorn et al., 2015; Schulz & Gershman, 2019). A range of tasks has been developed to examine exploration in humans (see Addicott, Pearson, Sweitzer, Barack, & Platt (2017) for a review). One of the most widely used tasks is the multi-armed-bandit task, which also allows an examination of how exploration behavior unfolds over longer periods. Here, participants make repeated choices between multiple (typically independent) choice options ("bandits") and observe a reward outcome following each choice. This task is assumed to require a balance between exploration and exploitation. Exploitation involves tracking each bandit's expected value and choosing the best. For exploration, on the other hand, at least two strategies are possible. First, exploration can be due to more or less stochastic selection of sub-optimal bandits, which in reinforcement learning can be modeled via ε-greedy or softmax choice rules (Daw et al., 2006; Schulz & Gershman, 2019; Sutton & Barto, 1998). Second, exploration can also entail a goal-directed component. In such models, agents track not only expected value, but also uncertainty. The probability to explore a bandit then scales with uncertainty about its expected value, thus maximizing information gain during exploration (Auer, Cesa-Biachini, & Fischer, 2002; Schulz & Gershman, 2019; Speekenbrink & Konstantinidis, 2015).

Exploration is supported by a bilateral fronto-parietal network including intra-parietal sulcus and fronto-polar cortex. (Badre, Doll, Long, & Frank, 2012; Daw et al., 2006; Raja Beharelle, Polania, Hare, & Ruff, 2015). While this network was initially characterized in the context of random exploration (Daw et al., 2006), recent evidence points towards a role of the fronto-polar cortex specifically in directed exploration (Badre et al., 2012; Boorman, Behrens, Woolrich, & Rushworth, 2009; Boorman, Behrens, & Rushworth, 2011; Zajkowski, Kossut, & Wilson, 2017). This is also supported by causal manipulations using transcranial magnetic stimulation (Zajkowski et al., 2017) and transcranial direct current stimulation (Raja Beharelle et al., 2015).

There is substantial evidence implicating the neurotransmitter dopamine (DA) in both exploration behavior and the pathophysiology of gambling disorder. Traditionally, DA has been associated with reward prediction error coding and exploitation (Beeler, Daw, Frazier, & Zhuang, 2010; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006). But both theoretical accounts (Beeler, 2012) and empirical data (Chakroun, Mathar, Wiehler, Ganzer, & Peters, 2019; Cinotti et al., 2019; Frank, Doll, Oas-Terpstra, & Moreno, 2009; Gershman & Tzovaras, 2018; Kayser, Mitchell, Weinstein, & Frank, 2014) suggest a role of DA in regulating the exploration-exploitation trade-off. DA also plays a role in problem gambling behavior. The most prominent empirical observation implicating DA in gambling comes from patients suffering from Parkinson's disease. These patients show higher rates of problem gambling behavior than the general population, which has been liked to pharmacological DA replacement therapy (Driver-Dunckley, Samanta, & Stacy, 2003; Voon et al., 2006). Gamblers may also exhibit increased pre-synaptic striatal DA levels (Boileau et al., 2014; van Holst et al., 2017), but this is controversially discussed (Majuri et al., 2017; Potenza, 2018).

Based on these observations, we hypothesized gambling disorder to be associated with a reduction in directed (uncertainty-based) exploration compared to healthy controls. In line with previous findings of a critical role of frontal pole regions (Daw et al., 2006; Raja Beharelle et al., 2015) and prefrontal dopamine (Frank et al., 2009) in exploration, we further hypothesized that reduced frontal pole recruitment would contribute to reduced exploration in gambling disorder. We addressed these issues by examining a group of gambling disorder participants and a group of healthy, matched controls using an established 4-armed bandit task during functional magnetic resonance imaging (fMRI, Daw et al., 2006). We used computational modeling to disentangle the effects of directed exploration and perseveration on choice behavior.

Methods

Sample

We investigated a sample of n=23 frequent gamblers (age mean [SD] = 25.91 [6.47], all male). Sixteen gamblers fulfilled four or more DSM-5 criteria of gambling addiction (mean [SD] = 6.31 [1.45], previously defined as pathological gamblers). Seven gamblers fulfilled one to three criteria (mean [SD] = 2.43 [0.77], previously defined as problem gamblers). All participants reported no other addiction except for nicotine (Fagerstrom test for nicotine dependence (FTND), (Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991)) score mean [SD] = 2.14 [2.58]). Current drug abstinence was verified via urine drug screening. All participants reported no history of other psychiatric or neurological disorder except depression. No participant was undergoing any psychiatric treatment. Current psychopathology was controlled using the Symptom Checklist 90 Revised (SCL-90-R) questionnaire (Schmitz et al., 2000) and depression symptoms were assessed via the Beck Depression Inventory-II (BDI-II, Osman, Kopper, Barrios, Gutierrez, & Bagge, 2004).

To further characterize gambling severity, we additionally conducted the German gambling questionnaire "Kurzfragebogen zum Glücksspielverhalten" (KFG, Petry, 1996), the German version of the South Oaks Gambling Screen (SOGS, Lesieur & Blume, 1987). We recruited n=19 healthy control participants, matched for age, gender, education, income, alcohol (Alcohol Use Disorders Identification Test, AUDIT, Saunders, Aasland, Babor, De la Fuente, & Grant, 1993) and nicotine consumption (Fagerstrom Test of Nicotine Dependence, FTND, Heatherton et al., 1991), see **Table 1** for sample characteristics and group comparisons. Participants were recruited via advertisements placed on local internet boards. All participants provided informed written consent prior to participation and the study procedure was approved by the local institutional review board (Hamburg Board of Physicians).

	GD mean	GD SD	HC mean	HC SD	t	df	p
Age	25.91	6.47	26.58	6.52	-0.33	38.42	0.74
School years	11.64	1.77	11.79	1.40	-0.31	39.93	0.76
Monthly income in EUR	1439.86	835.84	1015.74	588.05	1.92	39.10	0.06
FTND	2.14	2.58	2.21	2.18	-0.10	39.98	0.92
AUDIT	6.09	7.14	6.84	4.91	-0.40	38.85	0.69
DSM-5	5.13	2.22	0.11	0.32	10.72	23.07	<0.001
KFG	25.90	14.15	0.58	1.22	8.55	22.39	<0.001
SOGS	8.64	4.46	0.21	0.54	8.99	22.77	<0.001
BDI-II	15.41	11.41	8.47	8.46	2.26	39.60	0.03

Table 1. *Summary of demographics and group matching statistics. FTND: Fagerstrom Test of Nicotine Dependence, AUDIT: Alcohol Use Disorders Identification Test, KFG: Kurzfragebogen zum Glücksspielverhalten, SOGS: South Oaks gambling screen, BDI-II: Beck Depression Inventory-II. GD: Gambling disorder. HC: Healthy controls.*

Task and Procedure

Participants completed two sessions of testing on separate days. The first session included all questionnaires and an assessment of the spontaneous eye-blink rate, that was published previously (Mathar, Wiehler,

Chakroun, Goltz, & Peters, 2018). The second session started with a behavioral training session of the task, followed by functional and structural magnetic resonance imaging (MRI). During fMRI, participants first completed the bandit task reported here. Subsequently, they performed an additional task in the scanner that will be reported elsewhere.

We used a 4-armed bandit task, as described previously (Daw et al., 2006). We applied the same task as in the original publication, with the exception that we replaced the original slot machine images for each bandit with colored boxes (see **Figure 1A**). On each trial, participants selected one of the four bandits. Following the selection of a bandit, they received a payout between 0 and 100 points for the chosen bandit, which was added to a total score. The points that could be won on each trial were determined by Gaussian random walks, leading to payouts fluctuating slowly throughout the experiment (see **Figure 1B**, for mathematical details see Daw et al. (2006)). Participants completed 300 trials in total that were split into four blocks separated by short breaks. We instructed participants to gain as many points as possible during the experiment. Reimbursement was a fixed baseline amount plus a bonus that depended on the number of points won in the bandit task. In total, participants received between 70 and 100 Euros for participation.

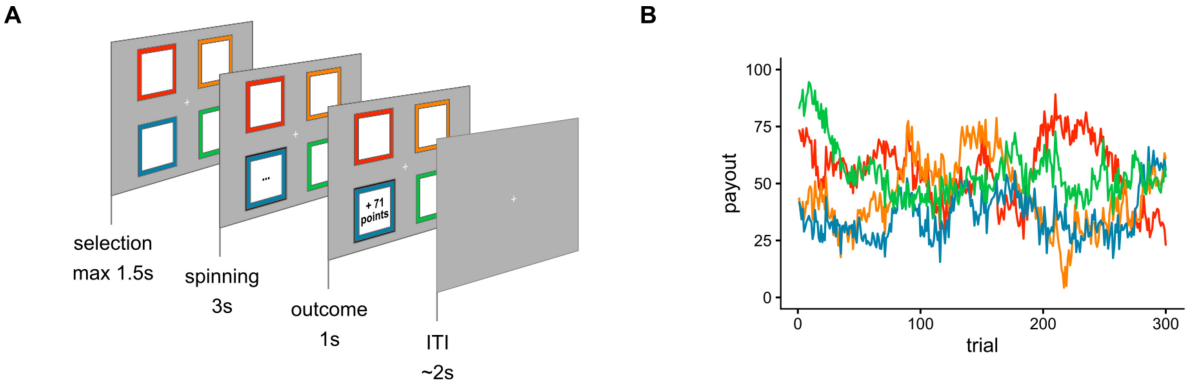

Figure 1. A: *One trial of the bandit task. On each trial, participants choose between four bandits on the screen and received a payout in reward points.* ***B****: Payouts fluctuated across the 300 trials of the experiment according to Gaussian random walks. Colors correspond to bandits in A. Here, one example set of random walks is shown.*

Computational modeling

To quantify exploration behavior, participants' choices were fitted with several reinforcement learning models of varying complexity. We first implemented a Q-learning model (Sutton & Barto, 1998). Here, participants update the expected value (Q-value) of the *i*th bandit on trial *t* via a prediction error δ_t:

$$Q_{i,t+1} = Q_{i,t} + \alpha\delta_t \tag{1}$$

with

$$\delta_t = r_t - Q_{i,t} \tag{2}$$

Here, Q is the expected value of the ith bandit on trial t, α is a learning rate, that determines the proportion of the prediction error δ_t that is used for the value update, and r_t is the reward outcome on trial t. In this model, unchosen bandits are not updated but retain their previous Q-values.

Q-values are transformed into action probabilities, using a softmax choice rule:

$$p(c_t = i) = \frac{exp(\beta Q_{i,t})}{\sum_j exp(\beta Q_{j,t})} \tag{3}$$

Here, p is the probability of choice c_t of bandit i in trial t, given the estimated values Q from equation 1 for all j bandits. β denotes an inverse temperature parameter, that models choice stochasticity: For greater values of β, choices become more dependent on the learned Q-values. Conversely, as β approaches 0, choices become more random. In this model, β controls the exploration-exploitation trade-off such that for higher values of β, exploitation dominates, whereas exploration increases as β approaches 0. Note, however, that this model does not incorporate uncertainty about Q-values, as only mean Q-values are tracked.

We next examined the Bayesian learner model (Kalman filter) that was also applied by Daw et al. (2006). This model assumes that participants use a representation of the Gaussian random walks that constitute the task's payout structure. Thus, irrespective of the choice, mean and variance of each bandit i are updated on each trial t as follows:

$$\hat{\mu}_{i,t+1} = \lambda \hat{\mu}_{i,t} + (1 - \lambda)\theta \tag{4}$$

$$\hat{\sigma}^2_{i,t+1} = \lambda^2 \hat{\sigma}^2_{i,t} + \hat{\sigma}^2_d \tag{5}$$

Here, μ is the mean expected value, σ is the standard deviation of the expected value, λ is a decay rate (fixed to 0.9836), θ is the decay center (fixed to 50), and σ_d is the standard deviation of the diffusion noise (fixed to 2.8). Note that these equations are used to generate the Gaussian walks (see Daw et al. (2006)). That is, without sampling, each bandits' mean value slowly decayed towards θ, and standard deviations increased σ_d units per trial.

The bandit chosen on trial t (c_t) is additionally updated using a delta rule similar to equation (2):

$$\hat{\mu}_{c_t,t+1} = \hat{\mu}_{c_t,t} + k_t \delta_t \tag{6}$$

with

$$\delta_t = r_t - \hat{\mu}_{c_t,t} \tag{7}$$

and

$$k_t = \frac{\hat{\sigma}^2_{c_t,t}}{\hat{\sigma}^2_{c_t,t} + \hat{\sigma}^2_o} \tag{8}$$

Equation 6 is analogous to equation 1, with one important exception: While the Q-learning model assumes that the learning rate is constant, in the Kalman filter the model learning rate is uncertainty-dependent. The trial-wise learning rate κ_t (Kalman gain) depends on the current estimate of the uncertainty of the bandit that is sampled (as per Eq. 8) such that the mean expected value is more updated when bandits with higher uncertainty are sampled. Specifically, $\sigma_{ct,t}$ refers to the estimated uncertainty of the expected value of the chosen bandit, and σ_o is the observation standard deviation, that is, the variance of the normal distribution from which payouts are drawn (fixed to 4). The uncertainty of the expected value of the chosen bandit is then updated according to

$$\hat{\sigma}^2_{c_t,t+1} = (1 - k_t)\hat{\sigma}^2_{c_t,t} \tag{9}$$

Taken together, this model gives rise to the following intuitions: First, participants not only track the expected mean payoff (μ) but also the uncertainty about the expected mean payoff (σ). The mean expected value of unsampled bandits is gradually moving towards the decay center and uncertainty about the value increases. Sampling of a bandit leads to a reduction in uncertainty (Eq. 9) that is proportional to the uncertainty prior to sampling. Additionally, the bandit's mean value is updated via the prediction error (Eq. 7) weighted by the trial-wise learning rate (Eq. 8) such that sampling from uncertain (but not certain) bandits leads to substantial updating.

We next combined this algorithm for value updating with three different choice rules for action selection. First, we used a standard softmax model (see Eq. 3). Here, choices are only based on the mean value estimates of the bandits $\mu_{i,t}$, such that exploration occurs in inverse proportion to the softmax parameter β and the differences in value estimates:

$$p(c_t = i) = \frac{exp(\beta\hat{\mu}_{i,t})}{\sum_j exp(\beta\hat{\mu}_{j,t})} \tag{10}$$

Second, we added an "exploration bonus" parameter φ that scales a bandit's uncertainty $\sigma_{i,t}$ and adds this scaled uncertainty as a value bonus for each bandit, as first described by Daw et al. (2006). This term implements directed exploration so that choices are specifically biased towards uncertain bandits.

$$p(c_t = i) = \frac{exp(\beta[\hat{\mu}_{i,t} + \varphi\hat{\sigma}_{i,t}])}{\sum_j exp(\beta[\hat{\mu}_{j,t} + \varphi\hat{\sigma}_{j,t}])} \tag{11}$$

Following a similar logic, we next included a parameter ρ modeling choice perseveration. ρ models a value bonus for the bandit chosen on the previous trial:

$$\mathbf{1}_{c_{t-1}}(i) := \begin{cases} 1 \text{ if } i = c_{t-1} \\ 0 \text{ if } i \neq c_{t-1} \end{cases} \tag{12}$$

$$p(c_t = i) = \frac{exp(\beta[\hat{\mu}_{i,t} + \rho \mathbf{1}_{c_{t-1}}(i)])}{\sum_j exp(\beta[\hat{\mu}_{j,t} + \rho \mathbf{1}_{c_{t-1}}(j)])} \quad (13)$$

Finally, we set up a full model including both directed exploration (φ) and perseveration (ρ) terms:

$$p(c_t = i) = \frac{exp(\beta[\hat{\mu}_{i,t} + \varphi \hat{\sigma}_{i,t} + \rho \mathbf{1}_{c_{t-1}}(i)])}{\sum_j exp(\beta[\hat{\mu}_{j,t} + \varphi \hat{\sigma}_{j,t} + \rho \mathbf{1}_{c_{t-1}}(j)])} \quad (14)$$

In total, our model space consisted of five models: 1) a Q-learning model with softmax, 2) a Bayesian learner with softmax, 3) a Bayesian learner with softmax plus exploration bonus, 4) a Bayesian learner with softmax plus perseveration bonus and 5) a Bayesian learner with softmax plus exploration bonus plus perseveration bonus model. All models were fitted using hierarchical Bayesian estimation in Stan version 2.18.1 (Carpenter et al., 2017) with separate group-level normal distributions for gamblers and controls for each choice parameter (β, φ, and ρ), from which individual-participant parameters were drawn. We ran four chains with 5k warmup samples and retained 10k samples for analysis. Group-level priors for means were set to uniform distributions over sensible ranges. Group level priors for variance parameters were set to half-Cauchy with mode 0 and scale 3.

Model comparison was performed using the Watanabe-Aikine Information Criteria, WAIC (Vehtari, Gelman, & Gabry, 2017; Watanabe, 2010) where smaller values indicate a better fit. To examine group differences in the parameters of interest (β, φ, and ρ) we examined the posterior distributions of the group-level parameter means. Specifically, we report mean posterior group differences, standardized effect sizes for group differences and Bayes Factors testing for directional effects (Marsman & Wagenmakers, 2017; Pedersen, Frank, & Biele, 2017). Directional Bayes Factors (dBF) were computed as *dBF = i / 1-i* where *i* is the integral of the posterior distribution of the group difference from 0 to +∞, which we estimated via non-parametric density estimation.

fMRI setup

MRI data were collected with a Siemens Trio 3T system using a 32 channel head coil. Functional MRI (fMRI) was recorded in four blocks. Each volume consisted of 40 slices (2 x 2 x 2 mm in-plane resolution and 1-mm gap, repetition time = 2.47s, echo time 26ms). We tilted volumes by 30° from the anterior and posterior commissures connection line to avoid distortions in the frontal cortex (Deichmann, Gottfried, Hutton, & Turner, 2003). Participants viewed the screen via a head-coil mounted mirror. High-resolution T1 weighted structural images were acquired after functional scanning was completed.

fMRI preprocessing

MRI data preprocessing and analysis was done using SPM12 (Wellcome Department of Cognitive Neurology, London, United Kingdom). First, volumes were realigned and unwarped to account for head movement and distortion during scanning. Second, slice time correction to the onset of the middle slice was performed to account for the shifted acquisition time of slices within a volume. Third, structural images were co-registered to the functional images. Finally, all images were smoothed (8mm FWHM) and normalized to MNI-space using DARTEL tools and the VBM8 template.

fMRI analysis

On the first level, we used two General Linear Models (GLMs) to model task-evoked effects using the canonical hemodynamic response function implemented in SPM. GLM 1 included the following regressors: 1) trial onset, 2) trial onset modulated by a binary parametric modulator coding whether the trial was a random exploration trial, 3) trial onset modulated by a binary parametric modulator coding whether the trial was a directed exploration trial, 4) outcome onset, 5) outcome onset modulated by model-based prediction error, and 6) outcome onset modulated by model-based expected value of the chosen bandit. Based on the best-fitting computational model, trials were classified as exploitation, directed exploration or random exploration. *Exploitation* trials are choices of the bandit with the highest expected value or the bandit with the highest sum of expected value and perseveration bonus. *directed exploration* trials are choices of the bandit with the highest exploration bonus. Finally, *random exploration* are all trials that are neither exploitation nor directed exploration. Error trials with missing responses were modeled separately. GLM 2 included the following regressors: 1) trial onset, 2) outcome onset, and 3) outcome onset modulated by the number of points earned. Again, error trials were modeled separately. Group differences were assessed by taking the single-subject contrast images to a second-level random-effects analysis (two-sample t-test). Here we included z-scored covariates for depression (BDI-II score), alcohol consumption (AUDIT score), smoking (FTND score), and age.

Dynamic causal modeling

Dynamic causal modeling (DCM, Stephan et al., 2008) allows specifying all nodes and connections of a hypothesized network model of brain activity. First, we extracted the BOLD time course of regions of interest (ROIs). ROIs were defined by group-level analysis (see results section). Time courses were extracted from a 5mm sphere around the single subject peak within a mask of the thresholded group activation. Models were constructed in a way that includes all permutations of inputs while keeping connections constant between regions. We assumed no modulations of connections. See the results section and the supplementary materials for more details on the tested models.

Classification analysis

To predict the group membership of each participant based on the DCM parameters, we used an unbiased, leave-one-pair-out approach. Dues to the unequal sample sizes, we repeated the classification with all possible subgroups of gamblers to match both groups in sample size. Within each of these subgroups, we trained a linear support vector machine classifier (SVM, Chang & Lin, 2011, C=1) on all participants except one patient and one control. We computed the prediction accuracy based on the left-out pair. We repeated this for all possible pairs and averaged accuracies across left-out pairs and sub-sampled groups. We repeated this procedure 500 times with randomly shuffled labels to build a null-distribution which allows to assess the significance of the observed accuracy.

Results

Model free results

The group difference in the number of points earned was not significant (controls mean [SD] = 18277.85 [1554.04], gamblers mean [SD] = 18489.69 [1520.32], $t_{38.19}$ = -0.44, p = 0.66). Median response times trended to be shorter in gamblers (controls mean [SD] = 0.44s [0.06], patients mean [SD] = 0.40s [0.07], $t_{39.37}$ = 1.75, p = 0.09).

Model comparison

Next, we used model comparison based on the Widely-Applicable Information Criterion (WAIC, where lower values indicate a better fit) to examine the behavioral data for evidence of directed exploration and perseveration. Choice data were fit in a hierarchical Bayesian estimation approach (see methods section) that assumes that individual subject parameters are drawn from group-level Gaussian distributions. In both groups, the Bayesian learning model (Kalman Filter) with softmax, exploration bonus, and perseveration bonus model accounted for the data best (see **Figure 2**). The same model ranking was replicated when we re-analyzed the original behavioral data from the Daw et al. (2006) study using our hierarchical Bayesian estimation approach (see **Figure S1**).

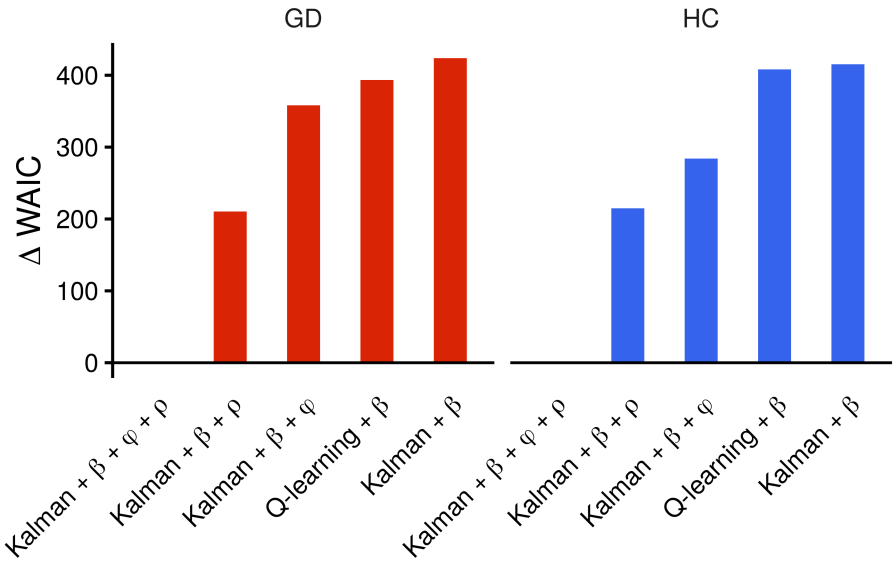

Figure 2. *Result of the model selection procedure based on WAIC (smaller WAIC = better model fit). In both gambling disorder patients (GD) and matched healthy controls (HC) a model based on a Kalman filter and including an uncertainty bonus (φ) and a perseveration bonus (ρ) wins.*

Parameters of the best-fitting model

Next, we analyzed the parameters of the best-fitting model in greater detail, focusing on the posterior distributions of the group means of choice stochasticity (softmax slope β), exploration bonus (directed, uncer-

tainty-based exploration φ), and perseveration bonus (ρ, see **Figure 3** and **Table 2**). There was evidence for a decrease in φ in the gamblers (see **Figure 3C, D** and **Table 2**) reflecting a decrease in directed exploration in the gamblers. Choice stochasticity β and perseveration ρ, on the other hand, were similar between groups such that the group difference distributions were in each case centered at zero (see **Figure 3A, B, E, F** and **Table 2**).

Parameter	**M_{diff}**	**Cohen's d**	**dBF**
β	0.002	0.16	1.22
φ	0.64	1.86	10.15
ρ	0.61	0.43	1.66

Table 2. *Summary of group differences for each choice parameter. M_{diff}: Non-standardized mean posterior group difference; Cohen's d: standardized mean posterior group difference, computed via the group mean posterior estimates of mean and variance; dBF: Bayes factor testing for directional effects (see methods section).*

We next explored whether individual differences in gambling addiction severity were associated with exploration behavior in the gamblers. As an index of addiction severity, we computed the mean z-score of SOGS and KFG scores. The correlation between addiction severity and single-participant φ parameters was not significant ($r=0.01$, $p = 0.95$).

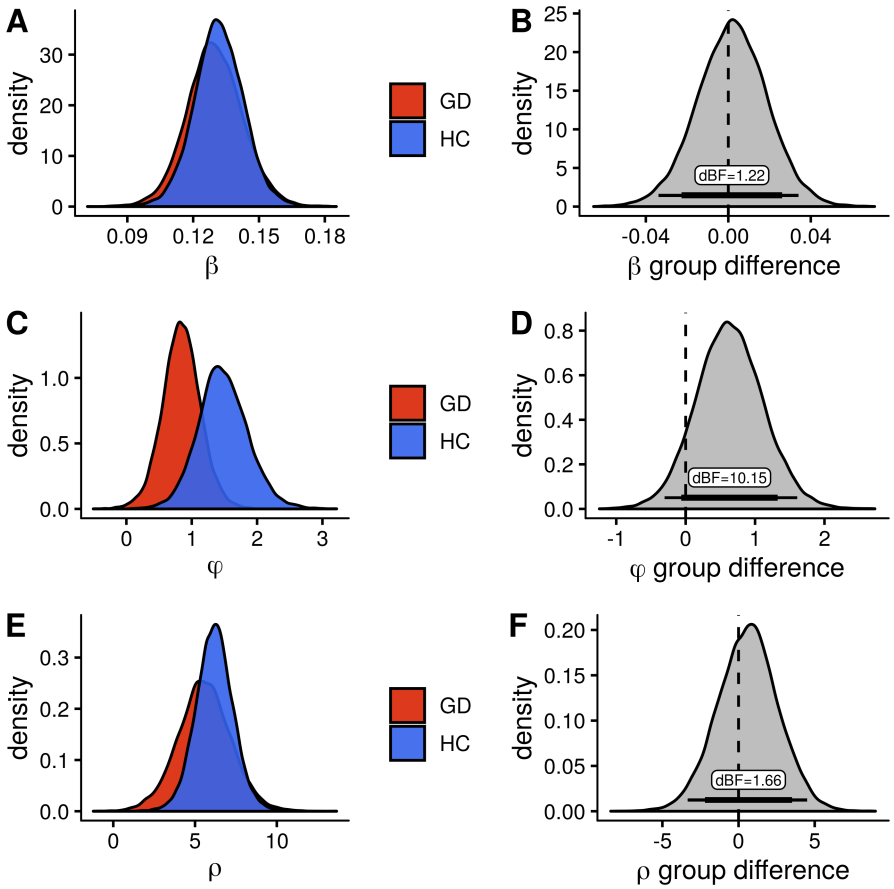

Figure 3. *Group parameters of the best-fitting model. GD: Gambling disorder, HC: Healthy controls.* ***A, C, E:*** *Posterior distribution of group-level parameters per group.* ***B, D, F:*** *Density of posterior distribution differences between groups. Bottom lines indicate the 85% and 95% highest density interval of the distribution. dBF: Directed Bayes factor, the proportion of the difference distribution above 0 over the proportion of the difference distribution below 0.*

fMRI

Group conjunctions

We first examined standard parametric and categorical contrasts, focusing on conjunction effects testing for consistent effects across groups. A first analysis examined outcome value, i.e. the parametric effect of points earned. Ventro-medial prefrontal cortex (vmPFC), ventral striatum (VS) and posterior cingulate cortex (PCC) parametrically tracked outcome value, in line with numerous previous studies and meta-analyses (Bartra, McGuire, & Kable, 2013; Clithero & Rangel, 2014; Daw et al., 2006, see Figure 4 and Table S1). Importantly, outcome value effects in these regions were observed across controls and patients, with no evidence for a group difference.

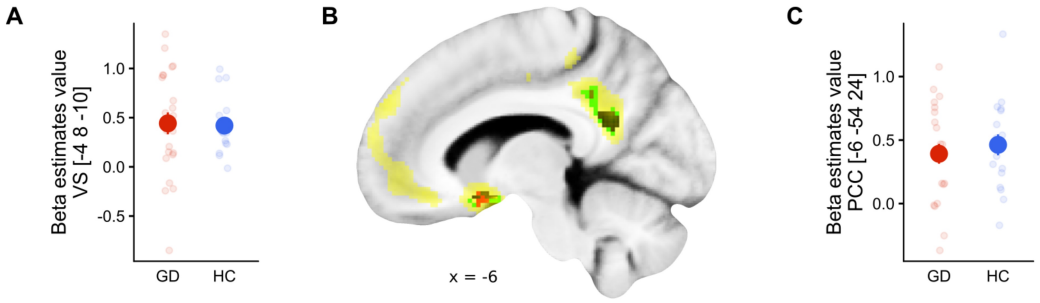

Figure 4. ***A****: Extracted beta estimates of each participant in the ventral striatum (VS).* ***B****: Display of the main effect value, yellow: p<0.001 uncorrected, red: whole-brain FWE corrected p<0.05, green: conjunction GD & HC p<0.001 uncorrected.* ***C****: Extracted beta estimates of each participant in the posterior cingulate cortex (PCC).*

We next computed model-based prediction errors for each trial based on the single-subject parameter estimates (medians of the posterior distribution) of the best-fitting hierarchical model. The ventral striatum coded prediction error in both groups (peak at x=-12, y=8, z=-16, main effect FWE corrected p<0.05, see **Figure S2)**, as previously described in healthy participants (Daw et al., 2006; Pessiglione et al., 2006).

Next, we analyzed exploration-related effects. Based on the best-fitting computational model, trials were classified as exploitation, directed exploration or random exploration (see Methods section). **Figure 5A and B** show the main effect of directed exploration with extensive effects in a fronto-parietal network, replicating previous findings using the same task (Daw et al., 2006, see also Table S2). Region-of-interest (ROI) analyses confirmed significant main effect clusters bilaterally in the frontal pole (10mm spheres at -27, 48, 4 and 27, 57, 6, (Daw et al., 2006)).

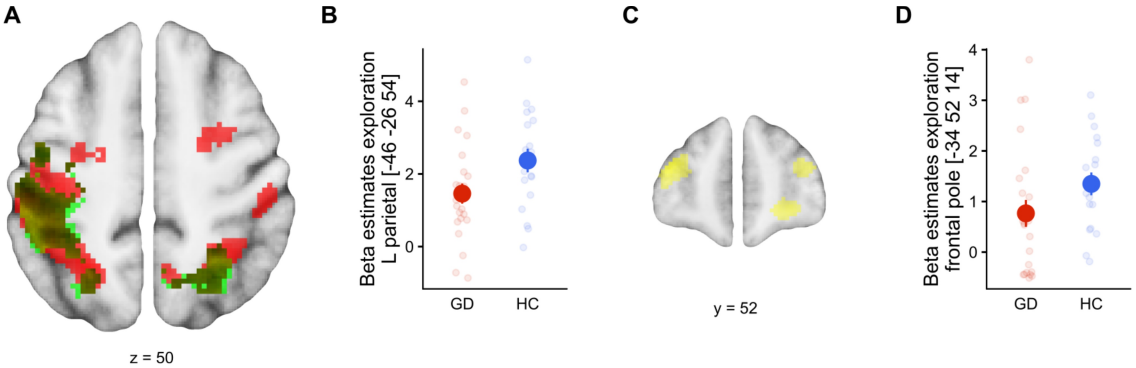

Figure 5. ***A****: Activations by directed exploration in the parietal cortex. red: Whole-brain FWE corrected p<0.05, green: Conjunction HC & GD, p<0.001.* ***B****: Beta extracts per participant from the left parietal cortex.* ***C****: Activations by directed exploration in the frontal pole, main effect, p<0.001.* ***D****: Beta extracts per participant from the left frontal pole.*

Group differences in exploration-related effects

We next tested our initial hypothesis of reduced frontal pole effects during directed exploration in the gamblers. Using an ROI approach, we checked previously identified bilateral frontal pole regions for evidence of group differences. Small volume FWE corrected analyses within 10mm spheres around peak activations of Daw et al. (2006) in the left (-27, 48, 4) and right (27, 57, 6) frontopolar cortex revealed no supra-threshold voxels ($p < 0.05$). We next performed an exploratory whole-brain analysis (at p<0.001 uncorrected) of group

differences in brain activity during directed exploration. Controls showed greater activation in parietal cortex (56, -26, 46, p<0.001 uncorrected) and in the substantia nigra / ventral tegmental area (SN/VTA, -12, -18, -10, p<0.001 uncorrected, **Figure 6** and Table S3). An exploration for effects of gambling severity on exploration-related brain activity in the gambling group revealed no supra-threshold effects even at an uncorrected threshold of p<0.001.

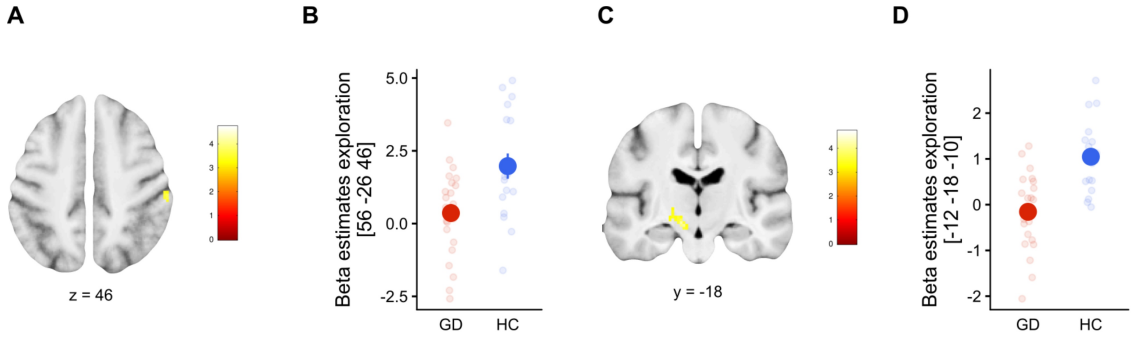

Figure 6: *A: Larger activations for healthy controls compared to gambling disorder in the parietal cortex, p<0.001. B: Beta extracts per participant from the peak voxel of A. C: Larger activations for healthy controls compared to gambling disorder in the SN/VTA, p<0.001. D: Beta extracts per participant from the peak voxel of C.*

Dynamic causal modeling and group differences in connectivity

Given the observation of largely overlapping exploration-related effects in fronto-parietal regions in the two groups, we next reasoned that group differences in network interactions might also contribute to the observed exploration deficit in the gambling group. Functional interactions within the task network might differ between groups and might thus be predictive of group status. To examine this possibility, we used dynamic causal modeling (DCM), a method to formally test and compare different causal models underlying observed task-related BOLD time courses. For each participant, we extracted the BOLD time-courses in three regions of interest (ROI) showing exploration-related effects at the group level: 1) left intraparietal sulcus (-46, -26, 54, main effect peak from the directed exploration contrast), 2) left frontopolar cortex (-34, 52, 14, main effect peak from the directed exploration contrast), and 3) left SN/VTA (-12, -18, -10, peak from the group comparison contrast directed exploration). We focused on the left hemisphere, as subcortical group differences were localized to the left SN/VTA.

As driving input, we used a binary regressor coding directed exploration trials (1) vs. other trials (0). All models included all reciprocal connections between the ROIs, but varied in the position of the input, ranging from no input to an input to all three ROIs (see **Figure S3** for an illustration of all models). Bayesian model selection (Stephan, Penny, Daunizeau, Moran, & Friston, 2009) revealed that the model with input confined to the parietal cortex accounted for the data best (expected probability = 0.59, exceedance probability = 0.99, see **Figure 7A** for a graphical depiction of the winning model and **Figure S4** for the model selection results). Further analysis then proceeded in two steps.

First, we extracted single-participant coupling and input-weight parameters and compared parameters between groups. Due to slightly different winning models between groups, we used Bayesian model averaging (Penny et al., 2010) which normalizes extracted parameters by the model evidence per participant. Parameters were compared between groups with two-sample t-tests and FDR correction for multiple comparisons. Several parameters showed a trend-level reduction in gamblers vs. controls (p<0.1, FDR corrected, frontal pole to frontal pole, input of explore in parietal cortex, input of explore in frontal pole, see **Figure S5**).

Second, we tested the hypothesis that the overall connectivity pattern contained information predictive of group membership (Brodersen et al., 2014). To this end, we used a support vector machine classifier to predict group membership based on all model parameters via a leave-one-pair-out, group-size balanced cross-validation scheme. The observed classification accuracy of 70.27% was significantly above chance level ($p<0.01$, permutation test with 500 randomly shuffled group labels, see **Figure 7B**). Thus, the DCM analyses confirmed that the pattern of functional network interactions contained information about group status, although several univariate analyses in these same ROIs did not reveal group differences.

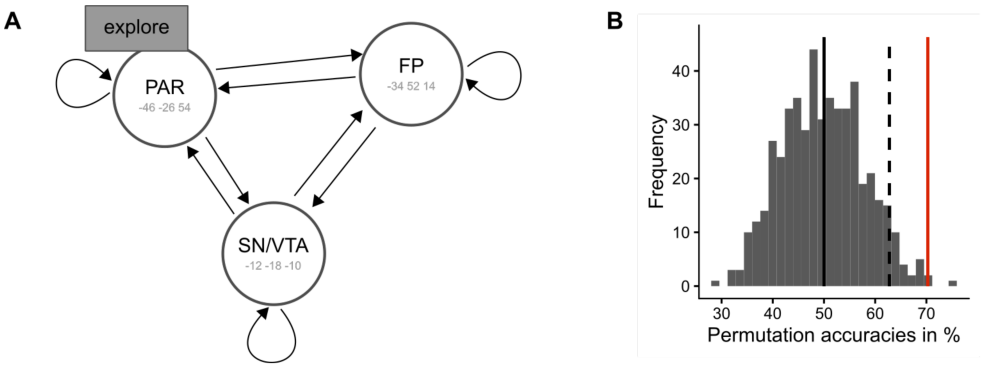

***Figure 7. A**: Illustration of the DCM winning in the model selection. explore: input of the explore regressor in parietal cortex, PAR: Parietal cortex, SN/VTA: substantia nigra / ventral tegmental area, FP: Fronto-polar cortex. The grey box indicates the driving input. **B**: Accuracy of the group classification based on the DCM parameters of the winning model (A). The histogram shows the H0 distribution of accuracy, created by classifications with randomly permuted labels. The dashed line is indicating the 0.95 quantile of the distribution, the red line is indicating the observed accuracy of 70.27%.*

Discussion

Here we used a combination of computational modeling and fMRI in the context of a well-established reinforcement learning task (four-armed restless bandit) to investigate reward exploration in gambling disorder. Computational modeling revealed that gamblers showed reduced directed exploration, whereas no group differences in perseveration were observed. FMRI revealed no significant group differences in the representation of basic task variables such as outcome value and reward prediction error. An exploratory analysis, however, revealed reduced activity during directed exploration in SN/VTA in gamblers. Dynamic causal modeling then showed that coupling among regions of an exploration network including parietal cortex, frontal pole, and SN/VTA was diagnostic of group membership (gamblers vs. controls).

Balancing exploration and exploitation is essential for reward maximization in volatile environments. Previous work showed that agents take the uncertainty of option values into account when making choices, that is, they show directed exploration (Schulz & Gershman, 2019). Our model comparison supported this. In both groups, the data were best accounted for by a Bayesian model that includes an exploration bonus term. However, in earlier models, estimates of exploration might have been confounded by choice perseveration (Payzan-LeNestour & Bossaerts, 2012; Wilson, Geana, White, Ludvig, & Cohen, 2014). If perseveration behavior is not accounted for in the model, this might lead to an underestimation of directed exploration. The reason is that perseveration-related variance can be misattributed as an exploration penalty, thereby increasing the proportion of participants showing negative φ parameters. This is particularly important in the context of clinical groups that are known to show increased perseveration (van Timmeren et al., 2018). We addressed this issue by extending existing models of exploration with an additional perseveration bonus term, such that final estimates of directed exploration are unconfounded by potential group differences in perseveration (Chakroun et al., 2019). Indeed, the full model including both directed exploration and perseveration terms accounted for the data best in both groups. Importantly, we replicated this model ranking in a re-analysis of the behavioral data from Daw et al. (2006, see supplemental information).

In the light of previous findings of reduced behavioral flexibility in gambling disorder, we hypothesized gamblers to show a specific reduction in directed exploration. This was supported by an examination of the posterior distributions of group-level model parameters. While both perseveration bonus parameter (ρ) and the random exploration / choice stochasticity β were very similar between groups, the directed exploration φ was substantially reduced in gamblers vs. controls (Cohen's d = 1.86). Alterations in reward-based learning and decision-making are well described in gambling disorder (Wiehler & Peters, 2015) and persistent gambling in the face of accumulating losses is a hallmark feature of the disorder. Through computational modeling, we disentangled perseveration and exploration accounts of flexibility impairments during reinforcement learning in gamblers and identified reduced directed exploration as the primary source of reduced exploration in gamblers. This finding might reflect an interference of maladaptive beliefs about environmental regularities and/or sources of influence on stochastic processes in gamblers with a "normal" or "natural" tendency to employ directed exploration. This reduction in directed exploration was not modulated by addiction severity, pointing towards a potential application of such paradigms in problem gamblers that show sub-clinical scores in traditional questionnaires, but might still be at risk to develop a more severe gambling disorder. Since gambling-related cognitive distortions can improve under cognitive-behavioral psychotherapy (Casey et al., 2017), this may also have the potential to attenuate such maladaptive interference processes.

While perseveration had a similar impact on decision-making in gamblers and controls in the restless bandit task, this does not rule out that perseveration might contribute to impairments in other tasks in gambling disorder. During reversal learning tasks, gamblers take more trials than controls to adapt to contingency changes (Boog et al., 2014; de Ruiter et al., 2009). Here, a continuous exploration of the alternative option could support a timely detection of reversals. Yet, increased perseveration could also underlie the reduced performance

of gamblers during reversal learning. However, after multiple reversals, the increased volatility of the environment (Behrens, Woolrich, Walton, & Rushworth, 2007) might drive participants to show more directed exploration, allowing a quicker detection of reversals. Reduced exploration could finally also apply to cognitive flexibility tasks like the Wisconsin Card Sorting Task, where participants are aware of frequent rule changes and could actively explore the possibility of a rule change (van Timmeren et al., 2018).

At the neural level, we found that basic task parameters were similarly represented in both groups. Value effects were localized in a well-characterized network encompassing vmPFC, ventral striatum and posterior cingulate cortex with no evidence for group differences, in line with previous meta-analysis (Bartra et al., 2013; Clithero & Rangel, 2014). Likewise, striatal prediction error signals were similar between groups. Again, this replicates findings in controls (McClure, Berns, & Montague, 2003; Pessiglione et al., 2006), and shows relatively intact prediction error signaling in gamblers. However, the nature of reward signals in gambling addiction remains an issue of considerable debate and inconsistency (Balodis et al., 2012; Clark et al., 2019; Leyton & Vezina, 2012; Miedl et al., 2014; Van Holst, Veltman, Van Den Brink, & Goudriaan, 2012). These inconsistencies might be due to specific differences in the implementation and/or analysis of the different tasks (e.g. anticipation vs. outcome processing, gain vs. loss domain, presence vs. absence of gambling-related cues or task-characteristics). However, our version of the four-armed bandit task includes neither gambling cues nor monetary reward cues or explicit probability information. These factors may have contributed to the null findings regarding the basic parametric effects of value and prediction error (Leyton & Vezina, 2012). Furthermore, few participants in our sample exhibited very high levels of addiction severity (compared to e.g. Miedl et al. (2012)). This might have precluded us from detecting more pronounced group differences in neural value and prediction error effects. We also did not observe correlations between addiction severity and behavioral and/or fMRI readouts. While this contrasts with some previous findings using other tasks (Miedl et al., 2012; Reuter et al., 2005; van Holst, Veltman, Büchel, van den Brink, & Goudriaan, 2012), overall such effects show considerable variability, both regarding behavior (Wiehler & Peters, 2015) and in reward-related imaging findings. Our study still included a considerable range of addiction severity (e.g. SOGS scores ranged from 3 to 17) suggesting that range restriction is an unlikely explanation for the lack of correlations. However, given the limited sample size typical of studies in addiction populations, statistical power is an additional concern also in the present study.

For the analysis of neural exploration effects, we extended previous approaches (Daw et al., 2006) by separating the neural effects of directed and random exploration via a model-based classification of trials. Again, overall effects were highly similar between groups and consistent with previous studies, such that directed exploration recruited a fronto-parietal network including frontal pole regions (Badre et al., 2012; Daw et al., 2006; Raja Beharelle et al., 2015). Importantly, our initial hypothesis that the reduction in directed exploration in gamblers was due to a down-regulation of frontal pole regions implicated in meta-control and exploration in decision-making (Shenhav, Cohen, & Botvinick, 2016) could not be confirmed. Although frontal pole and IPS effects of directed exploration were numerically smaller in the gamblers (see Figure 5), neither group difference was significant. Frontal pole effects of directed exploration were of very similar magnitude and distribution in both groups. The fact that frontal pole effects showed a right-lateralization as previously reported for explorations (Daw et al., 2006; Zajkowski et al., 2017), increasing our confidence in the robustness of the overall exploration effects in the imaging data.

We next examined functional connectivity within this exploration-related network via dynamic causal modeling. This revealed that group membership could be decoded from the DCM coupling parameters with a significantly above chance accuracy of 70.27%. This observation supports the idea that network interactions might contain more information reflecting a participants' clinical status than univariate contrasts (Brodersen et al., 2014). However, we emphasize that the overall prediction accuracy, though significantly above chance, is still too low for potential clinical decision making.

Given that dopamine has been implicated in both the exploration/exploitation trade-off (Beeler, 2012; Frank et al., 2009; Gardner, Schoenbaum, & Gershman, 2018; Kayser et al., 2014) and gambling disorder (Boileau et al., 2014; Majuri et al., 2017; Potenza, 2018; van Holst et al., 2017; Voon et al., 2006), we additionally carried out an exploratory analysis of subcortical correlates of directed exploration. Indeed, exploration-related activation in the dopaminergic midbrain (SN/VTA) was attenuated in gamblers. This finding resonates nicely with a recent study that reported increased dopamine synthesis capacity in striatal regions in gamblers (van Holst et al., 2017), but see Potenza (2018) for a critical discussion. Given the reciprocal connectivity between striatum and SN/VTA (Haber & Knutson, 2010), increased striatal-midbrain feedback inhibition might be one mechanism underlying the attenuated exploration-related midbrain activity. However, further sources of influence are possible given the alterations in circuit dynamics that we observed in the DCM analysis. Our categorical analysis of exploration is limited to the propensity to explore was analyzed for each given trial. This approach might neglect exploration tendencies that accumulate over longer periods, e.g. in networks tracking overall uncertainty, and this might constitute an interesting avenue for future research.

In addition to the limitations already addressed in previous sections (sample size, the magnitude of addiction severity scores), additional limitations need to be addressed. We did not randomize the position of the bandits on the screen. This precludes us from disentangling subcomponents of perseveration – perseveration might be due to selecting the same bandit again or to a repetition of the same motor action. Future studies might benefit from additionally randomizing bandit position between trials. Furthermore, it is still unclear whether exploration measured using different tasks taps into the same construct (von Helversen, Mata, Samanez-Larkin, & Wilke, 2018). It would thus be interesting to examine other tasks that operationalize exploration somewhat differently (Frank et al., 2009; Wilson et al., 2014; Zajkowski et al., 2017). As in previous studies using the four-armed bandit task (Raja Beharelle et al., 2015), reduced directed exploration did not translate into a reduced payoff. This appears to be an interesting feature of this task: Participants regularly adopt a directed exploration strategy, but this does not necessarily lead to a maximization of the overall payoff. It also remains unclear whether reduced directed exploration constitutes a vulnerability factor or a consequence of continuous gambling. As with other decision-making impairments in gambling disorder, it would be interesting to see whether these effects are tied to the clinical development of patients (e.g. to the escalation of gambling behavior or treatment effects of cognitive-behavioral therapy) or whether they manifest as stable factors that increase the risk for the development of the disorder. To clarify this, the field would need to move towards more longitudinal approaches. Finally, a comparison to substance-based disorders would be of considerable interest, in particular given the overlap in terms of decision-making impairments.

Impairments in reward-based learning, decision-making and cognitive control are hallmarks of gambling disorder. Here we show using computational modeling that during reinforcement learning in a four-armed restless bandit task, gamblers' behavioral impairments were attributable to reductions in directed exploration rather than increased perseveration. We observed no significant differences in the neural representations of model-based expected value and reward prediction error in striatal and ventromedial prefrontal cortex regions in gamblers and controls. Fronto-parietal exploration-related activity was also similar between groups but coupling parameters from a dynamic causal model of an exploration-related network contained information of clinical status. Finally, an analysis of subcortical exploration-related group differences revealed reduced activity in the SN/VTA in gamblers, which complements accumulating evidence for dopaminergic dysfunctions associated with this disorder (Boileau et al., 2014; van Timmeren et al., 2018). Taken together, our findings highlight a computational mechanism underlying decision-making impairments in gambling disorder during reinforcement learning and lend further support to the idea that dopaminergic dysregulation as a contributing factor.

Acknowledgments

A.W. and J.P. designed research. A.W. performed research. A.W. and K.C. analyzed the data. A.W. and J.P. co-wrote the paper, and K.C. provided revisions. J.P. supervised the project. This work was funded by Deutsche Forschungsgemeinschaft (Grant PE1627/5-1 to J.P.). We thank Anica Bäuning for assistance with task programming. We thank Raymond Dolan and Nathaniel Daw for kindly making the behavioral data from their 2006 paper available for re-analysis.

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research. *Neuropsychopharmacology*, (May), 1–33. https://doi.org/10.1038/npp.2017.108

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)* (5th editio). Arlington, VA: American Psychiatric Association.

Auer, P., Cesa-Biachini, N., & Fischer, P. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, *47*, 235–256. https://doi.org/10.1023/A:1013689704352

Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, *73*(3), 595–607. https://doi.org/10.1016/j.neuron.2011.12.025

Balodis, I. M., Kober, H., Worhunsky, P. D., Stevens, M. C., Pearlson, G. D., Potenza, M. N., … Vezina, P. Attending to striatal ups and downs in addictions. , 72 Biological Psychiatry § (2012).

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Beeler, J. A. (2012). Thorndike's Law 2.0: Dopamine and the Regulation of Thrift. *Frontiers in Neuroscience*, *6*, 116. https://doi.org/10.3389/fnins.2012.00116

Beeler, J. A., Daw, N., Frazier, C. R. M., & Zhuang, X. (2010). Tonic Dopamine Modulates Exploitation of Reward Learning. *Frontiers in Behavioral Neuroscience*, *4*(November), 170. https://doi.org/10.3389/fnbeh.2010.00170

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Boileau, I., Payer, D., Chugani, B., Lobo, D. S. S., Houle, S., Wilson, A. A., … Zack, M. (2014). In vivo evidence for greater amphetamine-induced dopamine release in pathological gambling: a positron emission tomography study with [11C]-(+)-PHNO. *Molecular Psychiatry*, *19*(12), 1305–1313. https://doi.org/10.1038/mp.2013.163

Boog, M., Höppener, P., v. d. Wetering, B. J. M., Goudriaan, A. E., Boog, M. C., & Franken, I. H. A. (2014). Cognitive Inflexibility in Gamblers is Primarily Present in Reward-Related Decision Making. *Frontiers in Human Neuroscience*, *8*(August), 569. https://doi.org/10.3389/fnhum.2014.00569

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. *Neuron*, *62*(5), 733–743. https://doi.org/10.1016/j.neuron.2009.05.014

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a Neural Network centered on human lateral frontopolar cortex. *PLoS Biology*, *9*(6). https://doi.org/10.1371/journal.pbio.1001093

Brodersen, K. H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W. D., Buhmann, J. M., & Stephan, K. E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage: Clinical*, *4*, 98–111. https://doi.org/10.1016/j.nicl.2013.11.002

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), 551–555. https://doi.org/10.18637/jss.v076.i01

Casey, L. M., Oei, T. P. S., Raylu, N., Horrigan, K., Day, J., Ireland, M., & Clough, B. A. (2017). Internet-Based Delivery of Cognitive Behaviour Therapy Compared to Monitoring, Feedback and Support for Problem Gambling: A Randomised Controlled Trial. *Journal of Gambling Studies*, *33*(3), 993–1010. https://doi.org/10.1007/s10899-016-9666-y

Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F., & Peters, J. (2019). Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. *BioRxiv*, 706176. https://doi.org/10.1101/706176

Chang, C.-C., & Lin, C.-J. (2011). {LIBSVM}: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:1--27:27.

Cinotti, F., Fresno, V., Aklil, N., Coutureau, E., Girard, B., Marchand, A. R., & Khamassi, M. (2019). Dopamine blockade impairs the exploration-exploitation trade-off in rats. *Scientific Reports*, *9*(1), 6770. https://doi.org/10.1038/s41598-019-43245-z

Clark, L., Boileau, I., & Zack, M. (2019). Neuroimaging of reward mechanisms in Gambling disorder: an integrative review. *Molecular Psychiatry*, *24*(5), 674–693. https://doi.org/10.1038/s41380-018-0230-2

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, *9*(9), 1289–1302. https://doi.org/10.1093/scan/nst106

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

de Ruiter, M. B., Veltman, D. J., Goudriaan, A. E., Oosterlaan, J., Sjoerds, Z., & van den Brink, W. (2009). Response perseveration and ventral prefrontal sensitivity to reward and punishment in male problem gamblers and smokers. *Neuropsychopharmacology*, *34*(4), 1027–1038. https://doi.org/10.1038/npp.2008.175

Deichmann, R., Gottfried, J. a., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage*, *19*(2), 430–441. https://doi.org/10.1016/S1053-8119(03)00073-9

Driver-Dunckley, E., Samanta, J., & Stacy, M. (2003). Pathological gambling associated with dopamine agonist therapy in Parkinson's disease: *Neurology*, *61*(3), 422–423. https://doi.org/10.1212/01.WNL.0000076478.45005.EC

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*(8), 1062–1068. https://doi.org/10.1038/nn.2342

Gardner, M. P. H., Schoenbaum, G., & Gershman, S. J. (2018). Rethinking dopamine as generalized prediction error. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1891), 20181645. https://doi.org/10.1098/rspb.2018.1645

Gershman, S. J., & Tzovaras, B. G. (2018). Dopaminergic genes are associated with both directed and random exploration. *Neuropsychologia*, *120*, 97–104. https://doi.org/10.1016/j.neuropsychologia.2018.10.009

Goudriaan, A. E., Brink, W. Van Den, & Holst, R. J. Van. (2019). Gambling Disorder. In A. Heinz, N. Romanczuk-Seiferth, & M. N. Potenza (Eds.), *Gambling Disorder*. https://doi.org/10.1007/978-3-030-03060-5

Haber, S., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology Reviews*, *35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. (1991). The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *British Journal of Addiction*, *86*, 1119–1127. https://doi.org/10.1111/j.1360-0443.1991.tb01879.x

Kayser, A. S., Mitchell, J. M., Weinstein, D., & Frank, M. J. (2014). Dopamine, Locus of Control, and the Exploration-Exploitation Tradeoff. *Neuropsychopharmacology*, *40*(2), 454–462. https://doi.org/10.1038/npp.2014.193

Kessler, R. C., Hwang, I., LaBrie, R., Petukhova, M., Sampson, N. A., Winters, K. C., & Shaffer, H. J. (2008). DSM-IV pathological gambling in the National Comorbidity Survey Replication. *Psychological Medicine*, *38*(9), 1351–1360. https://doi.org/10.1017/S0033291708002900

Lesieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): A New instrument for the Identification of Pathological Gamblers. *The American Journal of Psychiatry*, *144*(9), 1184–1188.

Leyton, M., & Vezina, P. (2012). On cue: Striatal ups and downs in addictions. *Biological Psychiatry*, *72*(10), 10–12. https://doi.org/10.1016/j.biopsych.2012.04.036

Lorains, F. K., Cowlishaw, S., & Thomas, S. A. (2011). Prevalence of comorbid disorders in problem and pathological gambling: systematic review and meta-analysis of population surveys. *Addiction*, *106*(3), 490–498. https://doi.org/10.1111/j.1360-0443.2010.03300.x

Majuri, J., Joutsa, J., Johansson, J., Voon, V., Alakurtti, K., Parkkola, R., … Kaasinen, V. (2017). Dopamine and Opioid Neurotransmission in Behavioral Addictions: A Comparative PET Study in Pathological Gambling and Binge Eating. *Neuropsychopharmacology*, *42*(5), 1169–1177. https://doi.org/10.1038/npp.2016.265

Marsman, M., & Wagenmakers, E. J. (2017). Three Insights from a Bayesian Interpretation of the One-Sided P Value. *Educational and Psychological Measurement*, *77*(3), 529–539. https://doi.org/10.1177/0013164416669201

Mathar, D., Wiehler, A., Chakroun, K., Goltz, D., & Peters, J. (2018). A potential link between gambling addiction severity and central dopamine levels: Evidence from spontaneous eye blink rates. *Scientific Reports*, *8*(1), 13371. https://doi.org/10.1038/s41598-018-31531-1

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal Prediction Errors in a Passive Learning Task Activate Human Striatum. *Neuron*, *38*(2), 339–346. https://doi.org/10.1016/S0896-6273(03)00154-5

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. a., … Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3), 191–215. https://doi.org/10.1037/dec0000033

Miedl, S. F., Büchel, C., & Peters, J. (2014). Cue-Induced Craving Increases Impulsivity via Changes in Striatal Value Signals in Problem Gamblers. *Journal of Neuroscience*, *34*(13), 4750–4755. https://doi.org/10.1523/JNEUROSCI.5020-13.2014

Miedl, S. F., Peters, J., & Büchel, C. (2012). Altered Neural Reward Representations in Pathological Gamblers Revealed by Delay and Probability Discounting. *Archives of General Psychiatry*, *69*(2), 177–186. https://doi.org/10.1001/archgenpsychiatry.2011.1552

Osman, A., Kopper, B. A., Barrios, F., Gutierrez, P. M., & Bagge, C. L. (2004). Reliability and Validity of the Beck Depression Inventory--II With Adolescent Psychiatric Inpatients. *Psychological Assessment*, *16*(2), 120–132. https://doi.org/10.1037/1040-3590.16.2.120

Payzan-LeNestour, É., & Bossaerts, P. (2012). Do not Bet on the Unknown Versus Try to Find Out More: Estimation Uncertainty and "Unexpected Uncertainty" Both Modulate Exploration. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00150

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin and Review*, *24*(4), 1234–1251. https://doi.org/10.3758/s13423-016-1199-y

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., & Leff, A. P. (2010). Comparing Families of Dynamic Causal Models. *PLoS Computational Biology*, *6*(3), e1000709. https://doi.org/10.1371/journal.pcbi.1000709

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051

Petry, J. (1996). *Psychotherapie der Glücksspielsucht*. Weinheim, Germany: Beltz, Psychologie-Verlag-Union.

Potenza, M. N. (2018). Searching for Replicable Dopamine-Related Findings in Gambling Disorder. *Biological Psychiatry*, *83*(12), 984–986. https://doi.org/10.1016/j.biopsych.2018.04.011

Raja Beharelle, A., Polania, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial Stimulation over Frontopolar Cortex Elucidates the Choice Attributes and Neural Mechanisms Used to Resolve Exploration-Exploitation Trade-Offs. *Journal of Neuroscience*, *35*(43), 14544–14556. https://doi.org/10.1523/JNEUROSCI.2322-15.2015

Reuter, J., Raedler, T., Rose, M., Hand, I., Gläscher, J., & Büchel, C. (2005). Pathological gambling is linked to reduced activation of the mesolimbic reward system. *Nature Neuroscience*, *8*(2), 147–148. https://doi.org/10.1038/nn1378

Saunders, J. B., Aasland, O. G., Babor, T. F., De la Fuente, J. R., & Grant, M. (1993). Development of the Alocohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Dection of Persons with Harmful Alcohol Consumption - II. *Addiction*, *88*, 791–804.

Schmitz, N., Hartkamp, N., Kiuse, J., Franke, G. H., Reister, G., & Tress, W. (2000). The Symptom Check-List-90-R (SCL-90-R): a German validation study. *Quality of Life Research : An International Journal*

of Quality of Life Aspects of Treatment, Care and Rehabilitation, *9*(2), 185–193. https://doi.org/10.1023/A:100893192

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, *55*, 7–14. https://doi.org/10.1016/j.conb.2018.11.003

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*(10), 1286–1291. https://doi.org/10.1038/nn.4384

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*(2), 351–367. https://doi.org/10.1111/tops.12145

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E. M., Breakspear, M., & Friston, K. J. (2008). Nonlinear dynamic causal models for fMRI. *NeuroImage*, *42*(2), 649–662. https://doi.org/10.1016/j.neuroimage.2008.04.262

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Sutton, R. S., & Barto, A. G. (1998). *Introduction to Reinforcement Learning* (1st ed.). Cambridge, MA, USA: MIT Press.

van Holst, R. J., Sescousse, G., Janssen, L. K., Janssen, M., Berry, A. S., Jagust, W. J., & Cools, R. (2017). Increased Striatal Dopamine Synthesis Capacity in Gambling Addiction. *Biological Psychiatry*, 1–8. https://doi.org/10.1016/j.biopsych.2017.06.010

van Holst, R. J., Veltman, D. J., Büchel, C., van den Brink, W., & Goudriaan, A. E. (2012). Distorted Expectancy Coding in Problem Gambling: Is the Addictive in the Anticipation? *Biological Psychiatry*, *71*(8), 741–748. https://doi.org/10.1016/j.biopsych.2011.12.030

Van Holst, R. J., Veltman, D. J., Van Den Brink, W., & Goudriaan, A. E. (2012, November). Right on cue? Striatal reactivity in problem gamblers. *Biological Psychiatry*, Vol. 72, pp. e23-24. https://doi.org/10.1016/j.biopsych.2012.06.017

van Timmeren, T., Daams, J. G., van Holst, R. J., & Goudriaan, A. E. (2018). Compulsivity-related neurocognitive performance deficits in gambling disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *84*(November 2017), 204–217. https://doi.org/10.1016/j.neubiorev.2017.11.022

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

von Helversen, B., Mata, R., Samanez-Larkin, G. R., & Wilke, A. (2018). Foraging, exploration, or search? On the (lack of) convergent validity between three behavioral paradigms. *Evolutionary Behavioral Sciences*, *12*(3), 152–162. https://doi.org/10.1037/ebs0000121

Voon, V., Hassan, K., Zurowski, M., Duff-Canning, S., de Souza, M., Fox, S., … Miyasaki, J. (2006). Prospective prevalence of pathologic gambling and medication association in Parkinson disease. *Neurology*, *66*(11), 1750–1752. https://doi.org/10.1212/01.wnl.0000218206.20920.4d

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Wiehler, A., & Peters, J. (2015). Reward-based decision making in pathological gambling: The roles of risk and delay. *Neuroscience Research*, *90*, 3–14. https://doi.org/10.1016/j.neures.2014.09.008

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074–2081. https://doi.org/10.1037/a0038199

Zajkowski, W. K., Kossut, M., & Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *ELife*, *6*, 1–18. https://doi.org/10.7554/eLife.27430