Addiction: A striatal roller-coaster

On the neural and associative-learning mechanisms underlying gambling and alcohol use disorder

van Timmeren, T.

Publication date
2019

Document Version
Other version

License
Other

Goal-directed and habitual decision making under stress in Gambling Disorder

van Timmeren T
Piray P
Goudriaan AE*
van Holst RJ*

** shared last authorship*

In preparation

3

Abstract

Background: Theories have suggested that a transition from goal-directed to habitual decision making underlies the development of addicted behaviors. However, research into these mechanisms has so far been confounded by the neurotoxic effects of drugs. Gambling disorder is a behavioral addiction characterized by continued gambling despite the harmful consequences, and provides a means to investigate addiction without the harmful effects of drugs on the brain. Additionally, stress is a known factor that prompts habitual behavior and increases the risk for addiction and relapse. We therefore tested whether acute stress would differentially impact the balance between goal-directed 'model-based' and habitual 'model-free' control systems in GD patients compared to HCs, as measured with the two-step reinforcement learning task.

Methods: Using a within-subject design, 22 patients with gambling disorder and 20 healthy controls underwent stress induction and a control condition before they performed the two-step task during fMRI. The balance between model-based and model-free decision making was measured using computational modeling.

Results: GD patients showed intact goal-directed decision making, which remained similar to HCs after stress induction. Bayes factors provided substantial evidence against a difference between the groups or a group-by-stress interaction on the balance between model-based and model-free decision making.

Conclusions: These results challenge the notion that addiction is related with an increased reliance on habits, even under a stressful situation, per se; instead, they suggest that putative distortions in habitual decision making may be the result of specific neurotoxic effects of drugs.

Introduction

Addiction is commonly defined as a chronic, relapsing neurobiological disease characterized by compulsive addictive behaviors despite negative consequences. One prominent theory suggests that addiction can be understood as the consequence of a disruption in the balance between goal-directed and habitual behavior (Everitt and Robbins, 2005, 2015). While initially goal-directed, addictive behaviors become increasingly driven by habits during the course of addiction and eventually become compulsive. This transition is suggested to be represented by a neural shift from prefrontal to striatal control. However, both animal and human investigations of habit formation in addiction have so far only focused on goal-directed decision making in substance use disorders (Sjoerds et al., 2013; Voon et al., 2015b; Ersche et al., 2016; Sebold et al., 2017). This has precluded conclusions about whether the putative 'shift' is caused by the neurotoxic effects of drug use, a consequence of repetitive addictive behaviors or a vulnerability marker for the development of addiction.

Gambling disorder has recently been recategorized as the first behavioral addiction in the substance-related and addictive disorders section of the DSM-5, a decision largely taken because of the clinical, phenomenological and neurobiological overlap with substance use disorders (Petry et al., 2014). It is unclear, however, whether impairments in goal-directed control are present in a 'behavioural addiction' like gambling disorder, where there is no drug involved (van Timmeren et al., 2018a). Because GD is free of the neurotoxic effects that confound neurobiological research in SUDs, GD can serve as a model that isolates the core features of addiction from the physiological consequences of drug use (Verdejo-García et al., 2008). Thus, investigating goal-directed control in GD can answer the important question of whether decreased goal-directed control is observed in the absence of the neurotoxic effects of drugs on corticostriatal systems.

Additionally, acute and chronic stress – both well-known risk factors for the escalation of and relapse to addictive behaviors (Koob and Le Moal, 2008; Sinha, 2008) – have been shown to prompt increased reliance on habitual decision-making (Schwabe and Wolf, 2009, 2010; Radenbach et al., 2015), mediated through cortisol (Otto et al., 2013). Theoretically, diminished goal-directed control through stress could be a crucial mechanism for addiction. Especially during early abstinence, acute stress may increase the (already enhanced) reliance on habitual control in patients with addiction, causing relapse (Schwabe et al., 2011). Remarkably little is known, however, about the behavioral and cognitive processes involved in the effects of stress on addictive behavior.

Based on the above, we set out to test whether acute stress would differentially affect goal-directed decision making in GD patients compared to HCs. We used a two-step reinforcement learning task (Daw et al., 2011), which offers a computational approach to studying instrumental learning. Specifically, the task allows to distinguish two computational

systems that control instrumental actions: a "model-free" system and a "model-based" system, the computational analogues of, respectively, habitual and goal-directed control (Daw et al., 2005; Gläscher et al., 2010). Using a within-subject crossover design, we tested the effect of stress on the balance between model-free and model-based decision making in GD patients and HCs. We hypothesized that goal-directed control in GD patients would be further decreased under acute stress.

Materials & Methods

We recruited 31 HCs and 26 GDs, but 7 participants (4 HCs) were excluded due to technical failure in one of two sessions, and 7 HCs and 1 GD were excluded because performance on the task indicated a lack of motivation: they repeated their choices ('stay') on >90% of the trials in at least one of the sessions. Thus, all analyses were performed on data from 20 HCs and 22 GD patients. GD patients were recruited from a local addiction treatment center (Jellinek, Amsterdam) and included if they were recently diagnosed with and started therapy for GD, but were not obliged to abstain from gambling. All subjects underwent a structured psychiatric interview [Mini-International Neuropsychiatric Interview–Plus] (Sheehan et al., 1998), which further confirmed criteria for DSM-5 Gambling Disorder in the GD group, or the lack thereof in HCs. Exclusion criteria for all subjects included: lifetime history of bipolar disorder, anxiety disorder, obsessive-compulsive disorder or schizophrenia; past six-month history of major depressive episode; current or past-year substance use disorder; current psychiatric treatment (except treatment for GD in GD patients); the use of any psychotropic medication; positive alcohol breath test or urine screen for (meth)amphetamines, benzodiazepines, opioids, cocaine, ecstasy, PCP, methadone or cannabis; history or current treatment for neurological disorders; major physical disorders; brain trauma; exposure to neurotoxic factors; colorblindness; or any contraindications for MRI. One subject (GD patient) tested positive on THC use, but because marijuana use occurred once, seven days prior to participation, and there was no history of dependence, this subject was included for further analyses.

All subjects provided written informed consent before participation. The study was approved by the Ethical Review Board of the Academic Medical Center and procedures were in accordance with the Declaration of Helsinki. Participants were reimbursed with 100€ plus additional task earnings (50€ on average) for their participation.

Procedure

Participants were tested on 2 separate days approximately 1 week apart (mean=8.1, SD=3.8 days), with both sessions starting at approximately the same time (average starting time=14:20h; mean time between start of sessions=32 min; SD=35min). All subjects were

tested in the afternoon to minimize time-of-day cortisol effects (Schwabe et al., 2008), except for one subject who was tested twice in the morning. In one of the sessions, participants underwent a stress manipulation (see section below) before entering the fMRI scanner to perform the two-step task (Daw et al., 2011) and a structural T1 and DTI MRI scan. In the control session, participants were asked to emerge their hand in lukewarm water before performing the two-step task, followed by another task (van Timmeren et al., under review) and a resting-state fMRI scan (van Timmeren et al., 2018b). On both testing days, participants were instructed on and practiced the two-step task before undergoing the stress or control manipulation. The order of the two sessions (control and stress) was counterbalanced across subjects.

On day one, participants completed the MINI interview, the Fagerstrom Test for Nicotine Dependence [FTND] (Heatherton et al., 1991) and the Alcohol Use Disorders Identification Test [AUDIT] (Saunders et al., 1993). On the second day, we tested participants' verbal IQ (using the Dutch Adult Reading Test (Schmand et al., 1991) and working memory (using the digit span, part of the Wechsler Adult Intelligence Scale; Wechsler, 1981). The experience of gambling-related problems was assessed using the past-12-month Problem Gambling Severity Index [PGSI] (Ferris and Wynne, 2001b) and the Gamblers' Beliefs Questionnaire [GBQ] (Steenbergh et al., 2002). The GBQ contains 21 items (e.g. 'My choices or actions affect the game on which I am betting' or 'I am pretty accurate at predicting when a "win" will occur'), with higher scores reflecting more gambling-related distortions.

Stress induction

To induce acute psychosocial stress, subjects underwent the Socially Evaluated Cold-Pressor Test [SECPT], a well-validated method for stress induction (Schwabe et al., 2008). Participants were asked to immerse one hand into ice water (0°–2° C) and keep it there as long as possible – or until the experimenter told them to stop (after 2 minutes). During this procedure, participants looked into a video camera and were closely observed by a nonsupportive experimenter who made notes and was dressed in a white doctor's coat. Subsequently, participants were asked to perform a challenging arithmetic task (counting backward from 2059 in steps of 17) in front of the experimenter. In the control condition, warm water (34°–38° C) was used, no camera was present, the arithmetic task was simple (counting in steps of 10) and the experimenter was supportive and casually dressed. After the control or stress induction, participants were brought to the fMRI scanner. Subjects started the two-step task approximately 13 (+/- 5) minutes after the SECPT; salivary cortisol peaks 15-45 min after stress induction (Schwabe et al., 2008).

Stress measurements

Saliva samples were taken using Salivettes® (Sarstedt, Germany) to measure cortisol levels before (at -15 min) and after (three times: at +10, +60 and +80 minutes) stress induction (t=0). Participants were asked to chew a cotton swab for ~1 minute. After testing, the samples were frozen and preserved at −22 °C until they were transported to the Dresden LabService (Germany) for analysis. Cortisol levels were not normally distributed and log-transformed in all statistical tests (Petzold et al., 2010). As done previously (Otto et al., 2013), cortisol delta was calculated by subtracting cortisol levels at t0 (pre-SECPT) from the average of t1 and t2 (post-SECPT) for each subject and session. Additionally, subjects rated how unpleasant, stressful and difficult to sustain they had experienced the procedure on a 7-point Likert scale immediately after the SECPT or control manipulation. For correlations with task performance, we used the difference between the stress and control condition for physiological (delta cortisol) and subjective stress measures. To minimize the effects of menstrual cycle on cortisol response (Kirschbaum et al., 1999), women were tested in the luteal phase (first visit 15-19 days after last menstruation).

Two-step Markov decision task

Subjects completed 201 trials of the two-step Markov decision task (Daw et al., 2011), designed to distinguish between model-free and model-based learning strategies (Figure 1A). Each trial consists of two stages. In the first stage, participants chose between two abstract stimuli depicted on a grey background. This probabilistically led to one of two second-stage states, represented by different background colors and stimuli pairs. Subjects again made a choice between two options, which then lead to an outcome (20 cent reward or no reward). Critically, the transition from the first choice to the second stage was probabilistic: each choice usually (70%) leads to one of the two second-stage states ('common transition') but sometimes (30%) to the other state ('rare transition'). This feature enables the distinction between model-based (goal-directed) and model-free (habitual) decisions on a trial-by-trial level, because the two decision strategies make distinct predictions on choice behavior (Figure 1B). The second-stage reward probabilities slowly drifted over time according to Gaussian random walks (reflecting boundaries at 0.25 and 0.75), to motivate participants to adjust their choices and learn throughout the task. Participants explicitly learned the transition frequencies during the training phase using different stimuli. The task was programmed in MATLAB (The MathWorks, Inc., Natick, MA, United States) with Psychophysics Toolbox, as previously used by (Sebold et al., 2017).

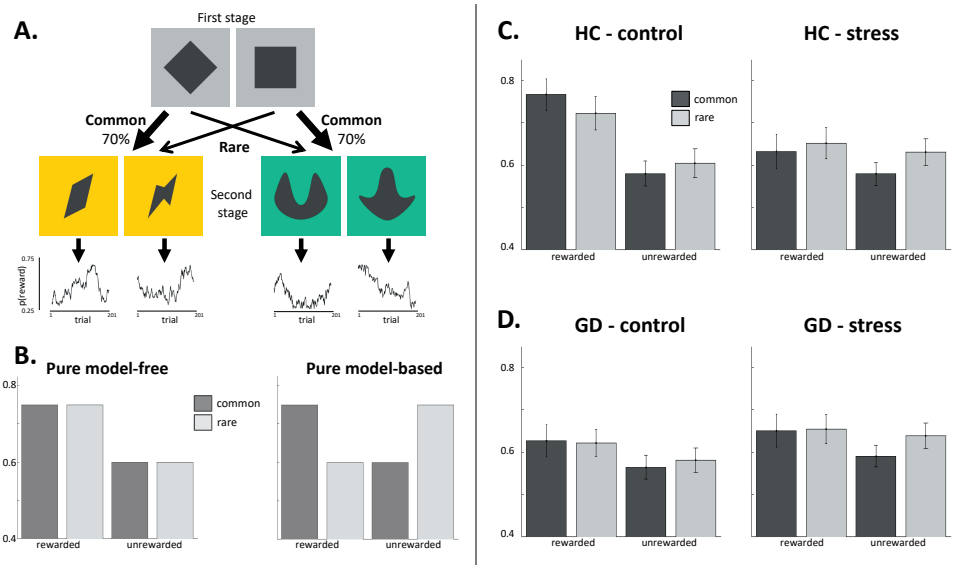

Figure 1. A. Schematic task, **B.** Model-free and model-based reinforcement learning strategies predict different responses based on the outcome of the previous trial. Model-free decisions are more likely to be repeated when the previous trial was rewarding, independent of the transition-type (common or rare). Model-based decisions, on the other hand, do take the transition probabilities into account and therefore an interaction between reward and transition-type is expected. **C. and D.** Across groups and sessions, a main effect of reward and an interaction between reward and transition-type was observed, indicating the presence of both model-free and model-based strategies. Additionally, there was a significant reward by group interaction, driven by lower stay probability after rewards in GD patients.

Data analysis

We investigated: 1) whether HCs and gambling disordered patients differed in the behavioral and neural signatures of model-free and model-based control; and 2) whether this balance would be differentially affected under acute stress in HCs and GDs. Statistical analysis were conducted in JASP software, version 0.9.0.0 (JASP Team, 2018), unless stated differently.

Behavioral analysis

As done previously, we focused on stay-switch behavior on the first stage choice of each trial to derive model-free and model-based strategies. First-stage choices were analyzed as a function of the previous trial's reward and transition-type. Because a model-free strategy disregards the structure of the task, a rewarding choice is more likely to be repeated and reflected by a main effect of reward on stay probability. Model-based choices, on the other hand, consider the transition probabilities from the first to the second

stage; therefore, receiving a reward after a rare transition increases the propensity to switch, reflected by an interaction between transition and reward on stay probability. Following previous work (Daw et al., 2011; Otto et al., 2013; Smittenaar et al., 2013; Piray et al., 2016), we analyzed the behavioral data in two complementary ways: by using a logistic regression model that captures model-free and model-based approaches by examining how the previous trial's outcome affects the next choice; and by using a full reinforcement learning model (the hybrid model from Daw et al., 2011) which allows choices to be influenced by the entire preceding history of outcomes.

For each subject, first-stage choices, encoded as binary stay/switch responses, were regressed against the factors reward, transition and stress and their interactions, resulting in a total of seven regressors and an intercept, reflecting the general tendency to stay (we used the glmfit routine in MATLAB, see also Table 2). Model-free and model-based control are represented, respectively, by the main effect of reward and the interaction effect between reward and transition. We then performed one-sample t-test on the individual coefficient estimates across all subjects and two-sample t-test to compare groups.

Additionally, data was fitted to the hybrid reinforcement learning model from Daw et al. (2011). This model contains seven parameters (see Figure 3), of which the weight parameter *w* captures the balance between model-free and model-based control. This weight parameter ranges from 0 (pure model-free) to 1 (pure model-based), with higher values of *w* reflecting a higher level of dependence on the model-based system. For model fitting, we used the 'computational and brain/behavior modeling' (CMB) toolbox (https://github.com/payampiray/cbm) in MATLAB. This toolbox offers a hierarchical and Bayesian inference framework for parameter estimation, which regularizes individual estimates according to group statistics through Hierarchical Bayesian Inference (HBI) to produce better individual estimates and permitting reliable group-level tests (for details see (Piray et al., 2018). The hybrid model with analytical gradient and Hessian, as originally implemented in (Piray et al., 2016), was used to facilitate optimization.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) was performed on a 3 Tesla, full-body Intera MRI scanner (Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel phased array SENSE radiofrequency (RF) receiver head coil. A high-resolution T1-weighted structural image was acquired for each participant (6.862 ms repetition time; 3.14 ms echo time; 8º flip angle; 1x1x1 mm voxel size; 236.679 x 180 x 256mm field of view; 212 x 212 matrix size; 150 slices; 1.2 mm slice thickness). Functional MRI scans were acquired using a T2*-weighted gradient multi-echo echoplanar imaging sequence (2375 ms repetition time; 9 / 26.4 / 43.8 ms echo times; 76º flip angle; 3x2.95x3 mm voxel size; 76 x 73 matrix size; 37 slices, acquired in interleaved order; 3mm slice thickness; 0.3mm slice gap). This sequence

was chosen for its improved blood oxygen level dependent (BOLD) sensitivity and lower susceptibility for artefacts, especially for ventral regions (Poser et al., 2006). The first three scans were discarded to allow T1 saturation to reach equilibrium.

fMRI analysis

Imaging data were preprocessed using SPM12 (Wellcome Trust Centre for Neuroimaging, London). Raw multi-echo data were combined as reported in van Timmeren et al. (van Timmeren, Zhutovsky, van Holst, & Goudriaan, 2018). In short, realignment parameters were estimated for the images acquired at the first echo time and consequently applied to images resulting from the two other echoes. The first thirty volumes, during which a fixation cross was shown, were used to calculate the optimal weighting of echo times for each voxel by applying a PAID-weight algorithm (Poser et al., 2006). The multi-echo fMRI data were then combined into single volumes using these weightings. Next, all functional images were slice-time corrected and co-registered with the high-resolution T1-weighted image using normalized mutual information. The high-resolution structural scan was segmented and used to normalize the slice-time corrected functional images. Finally, all functional images were smoothed with an 8mm isotropic full-width at half maximum (FWHM) Gaussian smoothing kernel.

For each participant, a first-level general linear model was constructed including the two sessions. First level analyses were conducted according to Daw et al. (2011). Model-free and model-based RPEs were derived from the computational model and the median across each group was used to generate a group-representative set of parameters. Model-free reward-prediction errors (RPEs) were used as parametric modulators at the second stage and outcome delivery onset to find BOLD activity that correlated with the model-free RPE signal. A second parametric regressor, defined as the difference between the model-free and model-based RPEs, captured BOLD activity related to model-based values. This regressor is only non-zero at the second-stage onset; to prevent the effect from being driven by the outcome delivery phase, we mean-corrected the regressor for each subject and session and included a nuisance regressor at the time of outcome onset (see Supplemental material for Daw et al. 2011). Additional nuisance regressors were included to capture first stage onset and movement (entered as the 6 separate regressors). A high-pass filter (128-s cutoff) was used to remove low frequency drifts and regressors were convolved with the canonical hemodynamic response function. Four first-level contrast images were constructed capturing the main effect of MF and MB RPE and their interaction with stress. These single-subject contrast images were then entered into second-level random-effects analysis, comparing within-group activation (one-sample t tests) and between-group differences (two-sample t tests). In line with Daw et al (2011), the model-based effect was captured by a covariate which included individual *w* values at second-level.

Results

Sample characteristics

Demographics and clinical information are presented in Table 1. Groups were matched for age, handedness, education, IQ and alcohol use (AUDIT). Initially, groups were also matched on gender, but after exclusions the remaining GD group contained significantly more males than the HC group ($p<0.02$). The number of GD subjects with nicotine dependence was significantly higher than in the HC group ($p=0.04$).

Table 1. Sample characteristics

	GD (n=22) Mean (SD)	HC (n=20) Mean (SD)	*p* value
Age, years	33.3 (12.7)	32.2 (13.8)	0.79
Males / females	18 / 4	9 / 11	**0.01**[a]
Handedness: right / left	20 / 2	17 / 3	0.56[a]
Education, years	7.6 (2.6)	9.1 (4.3)	0.14
Smokers (%)	11 (52%)	3 (15%)	**0.04**[a]
IQ	87.8 (9.5)	89.5 (11.9)	0.63
AUDIT	5.8 (4.7)	3.1 (2.1)	0.07[b]
PGSI (12 months)	14.5 (5.1)	0.2 (0.4)	**<0.001**[b]
Weeks abstinent	17.3 (23.7)	-	-

Demographical & Clinical information GD patients and matched controls. GD, Gambling Disordered patients; HC, Healthy Controls; SD, Standard Deviation; IQ, Verbal Intelligence Quotient; AUDIT, Alcohol Use Disorders Identification Test; PGSI, Problem Gambling Severity Index; GBQ, Gamblers' Beliefs Questionnaire; [a]*p* value of chi-square test. [b]Non-normally distributed data analyzed using Mann-Whitney U

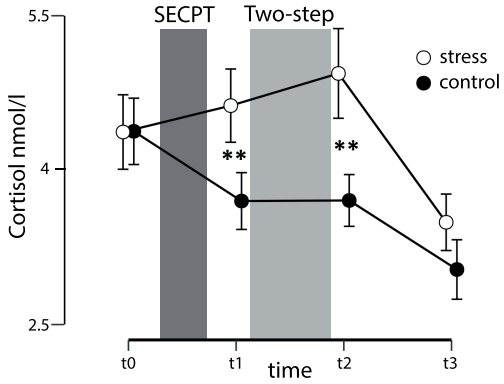

Figure 2. Salivary cortisol concentrations at different stages of the experiment. Cortisol was significantly increased after the SECPT compared to the control session, as measured both before (t1) and after (t2) performing the two-step task. There were no significant differences between HCs and GD patients. Data represent mean ± SEM across groups. **p<0.01.

Stress measures

A significant time-by-stress interaction indicated that cortisol was elevated in the stress compared to the control condition following the SECPT ($F_{3,114}$=2.9, p<0.02, η^2=0.08), indicating that stress induction was successful. Raw data are plotted in Figure 2. Moreover, the SECPT significantly elevated subjective stress levels, reflected by significantly higher ratings of unpleasantness (5.3 ± 1.6 vs 1.4 ± 0.9; t=12.9, p<0.001), stressfulness (4.5 ± 2.0 vs 1.3 ± 0.5; t=11.8, p<0.001) and difficulty to sustain (4.7 ± 2.0 vs 1.3 ± 1.0; t=10.0, p<0.001). No significant group differences were found.

Results stay-probability

Confirming the basic signatures of model-free and model-based strategies, respectively, the four-factor repeated-measures ANOVA (Group*stress*reward*transition-type) revealed a main effect of reward ($F_{1,40}$=28.6, p<0.001, η^2=0.38) and an interaction between reward and transition ($F_{1,40}$=6.3, p=0.01, η^2=0.13). Furthermore, a significant reward-by-group interaction was observed ($F_{1,40}$=6.6, p=0.01, η^2=0.09), driven by a lower stay probability after rewards in GD patients, indicating that GD patients repeated their choices less after a rewarding trial than HCs did (Figure 1C and 1D).

Results Logistic Regression

The logistic regression results mirrored findings from the stay probability analysis. A one-sampled t-test on the logistic regression coefficients indicated significant effects of reward (p<0.001, Cohen's d=0.91) and reward by transition (p=0.01, Cohen's d=0.43). The significant positive intercept indicates a general tendency to stay with the same choice regardless of transition and reward. A significant reward by stress interaction indicated that participants relied more on reward in the control condition than after stress induction (see Table 2). Group comparisons furthermore revealed that this effect was significantly different between groups (p=0.049), showing a significantly higher effect of stress on reward in HCs. Post-hoc tests revealed that there was a positive effect of stress on reward in HCs, indicating HCs were significantly more model-free after stress, but not in GD patients. The main effect of reward was also significantly lower in GD patients than in HCs (p=0.049). Table 3 provides the group-level coefficient values for the HC and GD groups and statistical comparison between them.

Table 2. The regressors included in the logistic regression analysis, indicating a main effect of reward (=model-free), an interaction between reward and transition (=model-based), an interaction between reward and stress and a main of the intercept, which represents the general tendency to repeat the same choice regardless of the other factors.

Logistic regression analysis of behavioral data (one-sampled t-tests)			
Effects	**Estimate (SEM)**	**t**	**p**
Reward	0.24 (0.04)	5.808	**< .001**
Transition	0.01 (0.03)	0.300	0.765
Reward X Transition	0.11 (0.04)	2.725	**0.009**
Reward X Stress	0.06 (0.03)	2.135	**0.039**
Transition X Stress	0.03 (0.02)	1.473	0.149
Reward X Transition X Stress	-0.01 (0.02)	-0.331	0.743
Stress	0.02 (0.05)	0.457	0.650
Intercept	0.63 (0.10)	6.365	**< .001**

Table 3. Comparing the regression coefficients between groups indicates that the effect of reward was weaker in GD patients than in controls. Furthermore, the groups differed on the interaction between reward and stress, driven by an effect of stress on reward in HCs but not in GD patients. Compare Figure 1C and 1D, which illustrate this difference.

Group comparison of logistic regression analysis (two-sampled t-tests)				
Effects	**HCs Estimate (SE)**	**GDs Estimate (SE)**	**t**	**p**
Reward	0.32 (0.07)	0.16 (0.04)	1.24	**0.049**
Transition	0.03 (0.04)	0.01 (0.03)	2.03	0.391
Reward X Transition	0.18 (0.07)	0.05 (0.04)	0.87	0.092
Reward X Stress	0.12 (0.04)	0.01 (0.04)	1.73	**0.049**
Transition X Stress	0.02 (0.04)	0.04 (0.03)	2.03	0.694
Reward X Transition X Stress	-0.04 (0.03)	0.02 (0.03)	-0.40	0.194
Stress	0.10 (0.07)	-0.05 (0.06)	-1.32	0.116
Intercept	0.76 (0.13)	0.52 (0.15)	1.61	0.224

Results computational modeling

Parameter estimates are plotted for both sessions and groups separately in Figure 3. A repeated measures ANOVA tested for an effect of group, stress or their interaction on the weighting parameter *w*. Contrary to our expectations, there was no significant difference between the two groups, nor did stress have a significant impact on the balance between model-based and model-free control (all p values>0.4). As this was the main question of the current study, we additionally quantified the evidence in favour of the null hypothesis against the evidence for the alternative hypothesis by means of the Bayes Factor BF_{01}. A

Bayesian repeated measures ANOVA provided substantial evidence for the absence of a group difference (BF_{01}=2.9), or the interaction between group and stress (BF_{01}=3.1).

We additionally compared all other parameters for differences between sessions, groups or their interaction. The only significant group difference was seen on β_2, which was lower in GD patients than HCs (main effect of group: $F_{1,39}$=4.2, p=0.04, η^2=0.10), indicating that GD patients were generally more random in their choices. Following previous work (Otto et al., 2013; Radenbach et al., 2015), we also investigated the relationship between delta cortisol (i.e. the difference between post minus pre-SECPT and post minus pre-control cortisol values) and the weight parameter (repeated measures ANOVA with $w_{control}$ and w_{stress} as within- and group as between-subject factor including delta cortisol as covariate) but failed to find any significant relationship (no main effect of delta cortisol, p=0.8, or an interaction with the weight parameter, p=0.18).

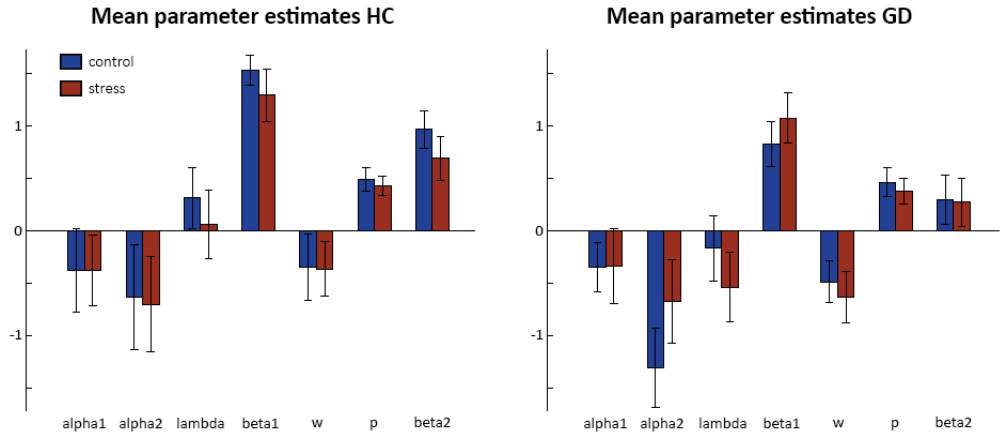

Figure 3. Mean estimates from the computational model for all seven parameters: learning rates for the first and second stage choices, α_1 and α_2; the eligibility trace parameter, λ; the weighting parameter w, which reflects balance between model-based and model-free values; repetition parameter ρ, reflecting perseveration; and two free inverse temperature parameters, β_1 and β_2, which reflect choice reliability. The first four parameters were logit-transformed and β_1 and β_2 were log-transformed (ρ was not transformed); thus, w=0 indicates an equal balance between the model-free and model-based values. Data represent mean ± SEM.

fMRI results

Across groups and conditions (control/stress), there was a main effect of model-free RPEs in regions previously associated with RPEs, including bilateral ventral striatum, caudate, putamen, ACC, pallidum, Insula (Figure 4A), but no significant correlates of model-based RPEs. No significant differences between groups were observed on the main effects of model-free or model-based RPE learning signals. Furthermore, no main effect of stress

was observed on model-free or model-based RPEs, nor did these effects differ between the groups.

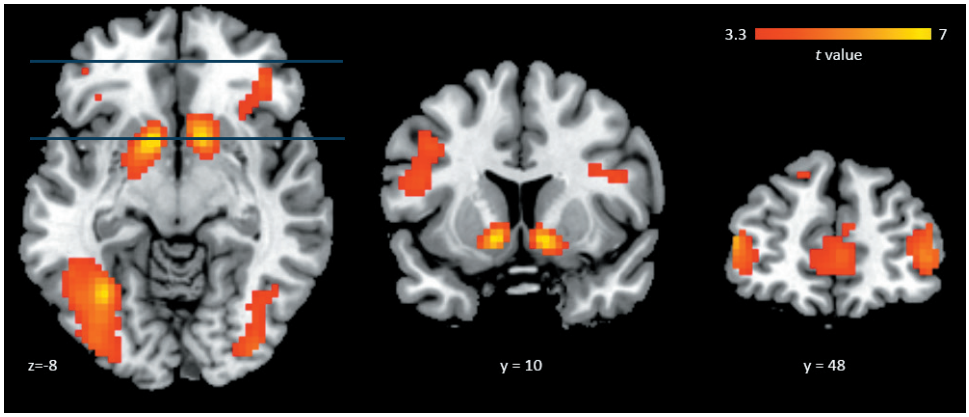

Figure 4. main effect of MF RPEs, with significant activations seen in several regions including the bilateral ventral and dorsal striatum and the mPFC ($p<0.05$, FWE-corrected). Displayed at $p<0.001$, uncorrected.

Discussion

This study tested the hypotheses that patients with gambling disorder show disrupted goal-directed 'model-based' and increased habitual 'model-free' decision making, and that stress would further shift this balance. Logistic regression analyses showed that the main effect of reward on the next choice (predicted by the model-free system) was significantly lower in GD patients, and that stress abolished this effect in HCs but in GD patients. However, when analyzed using the more comprehensive computational model, we found no evidence for differential model-free or model-based involvement in GD patients or under stress as an explanation for these group differences. In fact, there was substantial Bayesian evidence against a difference between the groups or a group-by-stress interaction on the balance between model-based and model-free decision making. Additionally, no differences in the neurobiological correlates of model-free and model-based reward prediction errors were observed.

Regarding the role of goal-directed learning deficits in addiction, a central but unresolved question relates to the role of changes induced by drugs: is impaired goal-directed control the consequence of prior drug use, of the repetitive addiction-related behavior itself, or a pre-existing vulnerability marker? According to the habit theory of addiction (Everitt and Robbins, 2005, 2015), progressively increased habit formation underlies the transition towards addiction; however, this theory does not explicitly distinguish between the effect

of drug exposure and addictive behavior itself. Using GD as a model for addiction without the confounding neurotoxic effects that characterize substance use disorders, our results indicate that goal-directed control is intact in the absence of drug abuse, as suggested by the weight parameter of the hybrid computational model. Studies assessing substance use disorders, however, have also reported mixed findings regarding the balance between mode-based and model-based control. One study reported increased reliance on model-free control in abstinent methamphetamine dependent subjects, but no difference with participants with alcohol use disorder compared to HCs (Voon et al., 2015b). Similarly, Sebold et al (2017) found no difference in model-free behavior in patients with alcohol use disorder compared to HC. Integrating these findings, one may conclude that drug-induced changes are largely responsible for goal-directed control deficits, with substance-specific differences.

A second question of our research pertained to goal-directed control under acute stress. Stress is an important factor in the onset and progression of addiction, and known to increase relapse risk (Sinha, 2007). In the case of GD, gambling may serve as a coping mechanism for acute or sustained stress (Coman et al., 1997; Raylu and Oei, 2002). Thus, stressful states may serve as gambling cues, increasing gambling behavior and triggering relapse. As stress has previously also been shown to increase habitual control (Schwabe and Wolf, 2009, 2011; Otto et al., 2013), we investigated whether acute stress would promote habitual decision making more in GD patients than in HCs. Contrary to our expectations, we found no evidence for such an interaction, suggesting that acute stress does not selectively shift the balance between goal-directed and habitual decision making in GD patients. However, although stress induction had a significant impact on salivary cortisol and subjective stress levels, this had no significant influence on goal-directed control in HCs. On closer look, the reported effects of stress on participant's performance on this task in previous studies have been subtle: using a between-subject design, Otto et al. (2013) found no main effect of condition (stress vs control), but instead a negative relation between individual cortisol stress response (independent of the stress manipulation) and model-based weight. Using a within-subject design, Radenbach et al. (2015) found a similar negative relationship between cortisol and model-based responding (again no main effect of stress), an effect that was even more pronounced with higher levels of chronic stress.

Despite the fact that there was no group difference and stress effect on the weight parameter *w*, which reflects the balance between model-free and model-based learning strategies, the logistic regression analysis indicated a significant group difference on the main effect of reward. This difference was driven by a lower beta coefficient in GD patients, indicating that the main effect of reward (reflective of model-free responding) was significantly lower in GD patients than in HCs. This finding implies that, although GD patients were more likely to repeat their actions when the previous trial was rewarding

(there was a significant main effect of reward), this probability was lower than in HCs. One explanation may be found in the more comprehensive computational modeling analysis, which showed significantly lower beta values in GD patients, indicating that choices were overall more random. Additionally, the logistic regression analysis showed a group difference on the interaction between reward and stress, which was driven by a significant effect of stress on reward in HC, but not GD patients: when stressed, HCs repeated their choices less after rewarding trials relative to the control session, whereas GD patients' behavior was not significantly impacted by stress. Again, an explanation for the absence for an effect of stress in GD patients may be that their responding was more random, thus not sensitive for the impact of stress. We can only speculate on why decisions by GD patients were less reliable, but one explanation may be GD patients did not fully comprehend the task structure. Individual variation in the balance between model-free and model-based performance is known to depend on personality factors such as IQ (Culbreth et al., 2016; Gillan et al., 2016), working memory (Otto et al., 2013; Schad et al., 2014) and impulsivity-levels (Deserno et al., 2015). Unmeasured differences in these factors between the groups may have contributed to suboptimal performance in GDs.

Several limitations need to be addressed. First, we had to exclude a relatively large number of subjects, in part due to the within-subject design which increases the chance of excluding participants due to something going wrong in one of the two sessions. We initially tested 26 GD patients and 31 HCs. Unfortunately, we had to exclude 8 subjects (4 from each group) due to technical errors in one of two sessions (a downside to the crossover design), and an additional 7 HCs and 1 GD because their choices on the task suggested unmotivated performance: they did not explore the task but instead repeated their choice on >90% of the trials, meaning they simply selected one choice as their favorite and stuck with it. This may reflect a lower motivation in HC participants to perform the (relatively complex) task. Since such behavior is uninformative regarding the balance between model-free and model-based behavior, these subjects were excluded. Second, after exclusions there were significantly more males in the GD group, which also contained significantly more smokers. Both gender and smoking are known to impact salivary cortisol stress responses (Kudielka and Kirschbaum, 2005); although there were no group differences on cortisol measures, these factors may still have impacted cortisol measurements and obscured possibly relevant effects, such as the relationship between the weight parameter *w* and cortisol values. The previous studies investigating this relationship tested only non-smoking males (Radenbach et al., 2015), or did not report these sample characteristics (Otto et al., 2013).

In sum, this study shows intact goal-directed decision making in GD patients, which remained similar to HCs after stress induction. Although these results initially seem surprising based on the habit theory of addiction (Everitt and Robbins, 2015), they appear to converge a larger body of recent findings in addicted populations. Goal-directed control

deficits, if already observed in addicted populations, at this point appear to be most likely the consequence of specific neurotoxic effects of some drugs.

Author contributions

T.v.T., R.J.v.H. and A.E.G. conceived and designed the study. T.v.T. acquired the data. T.v.T. carried out the data analyses, with support of P.P.; T.v.T. prepared the manuscript. All other authors provided critical revision of the manuscript for important intellectual content. All authors read, corrected and approved the final manuscript.

Supplementary Material

Supplementary Table 1. fMRI results across all participants (HC and GD groups).

Anatomical Region	L/R	X	Y	Z	k	FWE *p*	*t* value	Z
Fusiform gyrus	L	-33	-61	-10	498	<0.001	7.85	5.99
VS , Putamen, Caudate, Pallidum, hippocampus	L&R	-12	8	-7	477	<0.001	7.33	5.73
Middle Cingulate	R	3	-37	35	253	0.001	6.75	5.42
Parietal cortex	R	51	-40	47	154	0.002	6.55	5.3
Inferior Frontal	L	-48	44	8	165	0.006	6.02	4.99
Inferior Occipital	R	36	-85	-4	169	0.010	5.85	4.89
Inferior Parietal	L	-48	-49	47	308	0.016	5.67	4.78
Cerebellum	L	-39	-70	-37	86	*0.017	5.65	4.77
ACC, mPFC	R	9	41	17	195	*0.003	4.62	4.08

X, Y and Z coordinates are reported in MNI space. All p-values peak-level FWE-corrected, except *=cluster level FWE-corrected; k=cluster size; ACC= Anterior Cingulate Cortex; mPFC=medial prefrontal cortex.