Neurobiology of Disease

Behavioral and Neural Signatures of Reduced Updating of Alternative Options in Alcohol-Dependent Patients during Flexible Decision-Making

Andrea M.F. Reiter,[1,2,7] Lorenz Deserno,[1,3,4] Thomas Kallert,[5] Hans-Jochen Heinze,[1,4,6] Andreas Heinz,[3] and Florian Schlagenhauf[1,3]

[1]Max Planck Fellow Group "Cognitive and Affective Control of Behavioral Adaptation", Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany, [2]International Max Planck Research School on the Neuroscience of Communication (IMPRS NeuroCom), 04103 Leipzig, Germany, [3]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité - Universitätsmedizin Berlin, 10115 Berlin, Germany, [4]Department of Neurology, Otto-von-Guericke University, 39118 Magdeburg, Germany, [5]Soteria Clinic Leipzig, Helios Park-Klinikum Leipzig, 04289 Leipzig, Germany, [6]Department of Behavioral Neurology, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany and, [7]Chair of Lifespan Developmental Neuroscience, Department of Psychology, TU Dresden, 01602 Dresden, Germany

Addicted individuals continue substance use despite the knowledge of harmful consequences and often report having no choice but to consume. Computational psychiatry accounts have linked this clinical observation to difficulties in making flexible and goal-directed decisions in dynamic environments via consideration of potential alternative choices. To probe this in alcohol-dependent patients ($n = 43$) versus healthy volunteers ($n = 35$), human participants performed an anticorrelated decision-making task during functional neuroimaging. Via computational modeling, we investigated behavioral and neural signatures of inference regarding the alternative option. While healthy control subjects exploited the anticorrelated structure of the task to guide decision-making, alcohol-dependent patients were relatively better explained by a model-free strategy due to reduced inference on the alternative option after punishment. Whereas model-free prediction error signals were preserved, alcohol-dependent patients exhibited blunted medial prefrontal signatures of inference on the alternative option. This reduction was associated with patients' behavioral deficit in updating the alternative choice option and their obsessive-compulsive drinking habits. All results remained significant when adjusting for potential confounders (e.g., neuropsychological measures and gray matter density). A disturbed integration of alternative choice options implemented by the medial prefrontal cortex appears to be one important explanation for the puzzling question of why addicted individuals continue drug consumption despite negative consequences.

Significance Statement

In addiction, patients maintain substance use despite devastating consequences and often report having no choice but to consume. These clinical observations have been theoretically linked to disturbed mechanisms of inference, for example, to difficulties when learning statistical regularities of the environmental structure to guide decisions. Using computational modeling, we demonstrate disturbed inference on alternative choice options in alcohol addiction. Patients neglecting "what might have happened" was accompanied by blunted coding of inference regarding alternative choice options in the medial prefrontal cortex. An impaired integration of alternative choice options implemented by the medial prefrontal cortex might contribute to ongoing drug consumption in the face of evident negative consequences.

Introduction

A key characteristic of addictive disorders is that addicted individuals continue substance use despite evident harmful consequences. Addicted individuals regularly report having no choice but to consume. This suggests an impairment of integrating different choice options and their potential consequences. Thus,

Received Nov. 30, 2015; revised Aug. 7, 2016; accepted Aug. 14, 2016.

Author contributions: A.M.F.R., L.D., H.-J.H., A.H., and F.S. designed research; A.M.F.R., T.K., and F.S. performed research; A.M.F.R., L.D., and F.S. analyzed data; A.M.F.R., L.D., T.K., H.-J.H., A.H., and F.S. wrote the paper.

This study was supported by the Max Planck Society and by grants from the German Research Foundation awarded to F.S. (DFG SCHL1969/1-1, DFG SCHL 1969/2-2). We thank all of the patients who participated in this study. We also thank T. Dieterlen, K. Hudl, M. Kerkemeyer, R. Kratzer, L. Luettgau, C.D. Radenbach, T. Schmidt, C. Steffler, H. Teller, and T. Wilbertz for their assistance in recruitment and data acquisition. In addition, we thank H. Schmidt-Duderstedt for her help in designing the figures, E. Kelly for proofreading, and S. Valk for helpful comments on an earlier version of this manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Andrea M. F. Reiter, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany. E-mail: reiter@cbs.mpg.de.

DOI:10.1523/JNEUROSCI.4322-15.2016

 0270-6474/16/3610935-14$15.00/0

neglecting "what might have happened" may rigidly bias decision-making toward choice options that have been proven to be rewarding in the past (Chiu et al., 2008; Redish et al., 2008; Dayan, 2009).

Computational psychiatry accounts (Montague et al., 2012) have theoretically linked these maladaptive decision-making processes to disturbed mechanisms of inference (Huys et al., 2015), for example, difficulties learning the statistical regularities of the environmental structure to guide decisions. Deficits in cognitive flexibility are well known in patients experiencing addiction (Bechara and Damasio, 2002; Garavan and Stout, 2005; Ersche et al., 2011; Goldstein and Volkow, 2011). Thus, addiction has been theorized as one prime example of a breakdown of behavioral control in favor of simple and inflexible learning processes (Everitt and Robbins, 2005; Dayan, 2009; Lucantonio et al., 2012) with support from first behavioral studies (Sebold et al., 2014; Voon et al., 2015). One such example is model-free reinforcement learning (RL), where choice values are adjusted via learning from past rewards only. However, model-free RL neglects the environmental structure, for example the relation between chosen and unchosen options. Reversal learning is a well known paradigm challenging the individual to flexibly adapt behavior, and addicted individuals are impaired in such tasks (Izquierdo and Jentsch, 2012). However, in alcohol-dependent patients, parameters of model-free RL did not account for the observed deficit in flexible behavioral adaptation, and neural signatures of model-free RL did not differ between groups (Park et al., 2010; Deserno et al., 2015c). One potential explanation is that alcohol-dependent patients are specifically impaired in inference regarding the interdependencies of choice values (e.g., if one option is bad, the other one might be good), which might hamper alcohol-dependent individuals in flexibly adapting their behavior.

In the same vein, concurrent tracking of how different decision options relate to each other, thus, generalization about what might have happened, promotes flexible behavioral adaptation in healthy individuals (Hampton et al., 2006; Gläscher et al., 2009; Li and Daw, 2011; Schlagenhauf et al., 2014). In such reversal learning tasks, this depends on inference regarding the anticorrelated task structure, for example, when a drop in one decision value implicates a rise of the other value (Fig. 1*A*). Neural signatures of flexible behavioral adaptation and also model-free RL were previously found in a network consisting of ventral striatum as well as the medial and lateral prefrontal cortices (O'Doherty et al., 2004; Hampton et al., 2006; Daw et al., 2011; Deserno et al., 2015b). There is evidence that the medial prefrontal cortex (mPFC) is a key region in the concurrent tracking of choice values and thus enables flexible behavioral adaptation (Hampton et al., 2006). Here, we probe whether the modification of basic model-free RL with respect to the interdependencies of choice values, reflecting the anticorrelated environmental structure, is disturbed in alcohol addiction and whether this relates to the clinical feature of obsessive drinking.

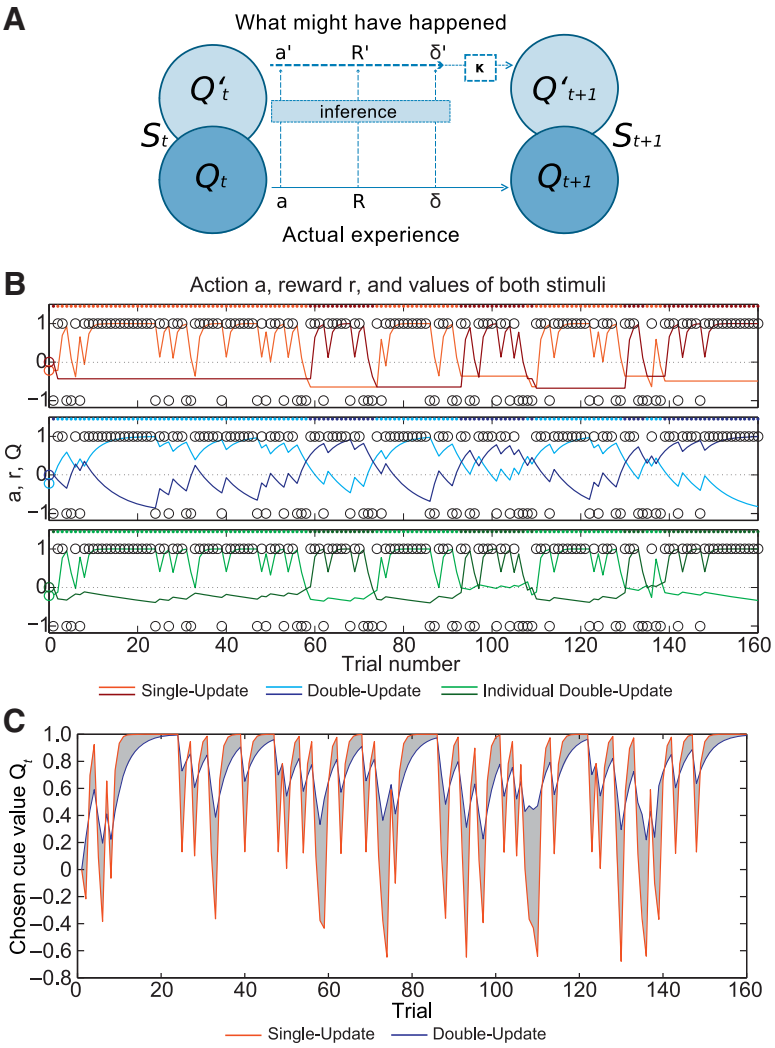

Figure 1. Schematic: parallel double-updating of chosen and unchosen choice values. ***A***, At time t, an agent in state S_t passes to a new state S_{t+1} by the action a, observing the outcome R, which leads to the reward prediction error δ as the difference between an expected and an actually gained reward. Accordingly, the agent updates the chosen value for the next trial, Q_{t+1}. Although not explicitly observed, the agent can conclude from the anticorrelated task structure what might have happened (R') if he had chosen an alternative action a', resulting in a fictive prediction error δ'. Thus, by inference on the anticorrelated task structure and parallel to updating chosen values, the agent additionally double-updates unchosen values Q'_{t+1}. Individuals might differ in their degree of inference on the environmental structure. The individual degree of double updating is therefore weighted by the parameter κ. ***B***, Trajectories of values of both stimuli as a function of κ (top: $\kappa = 0$, single-update model; middle: $\kappa = 1$, double-update model; bottom: $\kappa =$ free parameter, individually weighted double-update model) for one exemplary participant. Small colored dots in the upper edge of the figure indicate the chosen stimulus per trial, black circles indicate outcome per trial (1, reward; -1, punishment). The figure was generated by adapting plotting functions included in the HGF toolbox as part of the TNU Algorithms for Psychiatry Advancing Science (TAPAS Mathys et al., 2014). ***C***, Effect of inference, double-updating, on chosen values. For one exemplary participant, values of the respective chosen option are plotted per trial, as a function of the two alternative control strategies: pure single updating ($\kappa = 0$, neglecting what might have happened, red) vs pure-double-updating ($\kappa = 1$, full inference on the task structure, blue). Hence, the difference of both (here, highlighted in gray) represents an estimate of the degree of inference on the anticorrelated task structure. In our analysis of functional imaging data, we probe how this difference in choice values modulates the coding of the core teaching signal, the reward prediction error δ for chosen values.

Table 1. Sample characteristics

	Control subjects	Patients	Test statistic
Demographic characteristics			
Age (35/43)	42.00 ± 10.49	44.42 ± 10.21	$t = 1.03$, $p = .307$
Sex (male/female, 35/43)	25/10	34/9	$\chi^2 = 0.434$
Smokers (35/42)	16	33	$\chi^2 = 0.003$
Handedness (Edinburgh handedness scale, right/both/left, 35/39)	32/0/3	33/5/1	$\chi^2 = 0.521$
School leaving qualification (none/compulsory basic secondary schooling/intermediate school certificate/university entrance qualification, 35/41)	0/5/14/16	1/12/25/3	$\chi^2 = 0.001$
Total years of unemployment (35/41)	0.9 ± 1.58	4.54 ± 6.37	$t = 3.27$, $p = 0.002$
Neuropsychological measurements			
Reasoning (matrices, 35/41)	10.91 ± 4.00	6.71 ± 3.64	$t = 4.80$, $p < 0.001$
Working memory (backward digit span, 35/42)	7.49 ± 2.50	6.19 ± 2.00	$t = 2.54$, $p = 0.013$
Cognitive speed (TMT A, 35/42)	27.31 ± 14.44	38.82 ± 18.10	$t = -3.04$, $p = 0.003$
Complex attention (TMT B, 35/42)	62.84 ± 28.59	101.82 ± 79.52	$t = 2.75$, $p = 0.007$
Cognitive speed (DSST, 35/41)	79.91 ± 18.38	60.85 ± 16.14	$t = 4.81$, $p < 0.001$
Premorbid IQ (German vocabulary test, 35/41)	31.74 ± 3.38	24.20 ± 6.96	$t = 5.85$, $p < 0.001$
Barrat impulsiveness scale (35/42)	59.96 ± 10.03	65.81 ± 9.18	$t = 2.80$, $p = 0.007$
Clinical characteristics			
Alcohol units (month before participation/beginning of treatment, 35/38)	20.43 ± 21.67	301.61 ± 294.06	$t = 5.64$, $p < 0.001$
Obsessive-compulsive drinking scale (31/42)	3.65 ± 3.86	25.55 ± 9.78	$t = 11.80$, $p < 0.001$
Alcohol use disorder identification test (35/42)	4.26 ± 3.18	26.24 ± 8.72	$t = 14.14$, $p < 0.001$
Alcohol craving questionnaire (34/42)	1.3 ± 0.38	2.04 ± 0.88	$t = 4.42$, $p < 0.001$
Duration of dependence (years) (36)		14.64 ± 9.96	
Preceding detoxification treatments (35)		3.43 ± 3.99	
Beck depression inventory (33/41)	5.09 ± 6.32	13.59 ± 10.02	$t = 4.24$, $p < 0.001$

Data are reported as the mean ± SD, unless otherwise indicated. DSST, Digit symbol substitution test; TMT, trail making test.

To address this, we used functional magnetic resonance imaging (fMRI) during decision-making in a dynamic environment to examine flexible behavioral adaptation. Importantly, reward contingencies of different options were anticorrelated: whenever one stimulus was a good choice, the other one would be the worse choice, and vice versa. When confronted with options such as those in this task, individuals make choices based on decision values computed for the options at hand (Rangel et al., 2008). These can either be deduced by action–reward pairings or by inference on the anticorrelated reward probabilities (Hampton et al., 2006; Bromberg-Martin et al., 2010). We hypothesized that alcohol-dependent patients fail to integrate this inference, "what might have happened," into the value of the chosen options. To this end, we compared RL models that differ in updating the unchosen option. As a neural substrate, we predicted prefrontal signatures reflecting inference on alternative options to be reduced in alcohol-dependent patients.

Materials and Methods

Participants. Forty-three alcohol-dependent patients and 35 healthy participants were included. fMRI data were available for 35 healthy participants and 34 patients. Patients were recruited from an inpatient detoxification and rehabilitation program (Soteria Klinik Leipzig) and had abstained from alcohol for at least 8 d (range, 8–56 d; mean, 28.80 d; SD, 11.85 d). All patients were free of any psychotropic medication for at least four plasma half-lives except for one patient taking doxepin due to sleeping problems. All subjects underwent the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders* (DSM), fourth edition, Axis I Disorders (SCID-I; First et al., 2001) and patients additionally underwent a semi-structured interview on their individual addiction history. Alcohol dependence was diagnosed in all patients according to DSM, fifth edition, and International Statistical Classification of Diseases and Related Health Problems, 10th revision. Alcohol-dependent patients did not meet the criteria of any current comorbid psychiatric disorder. Included control participants did not report any current nor past psychiatric disorder (SCID-I). See Table 1 for demographic, neuropsychological, and clinical characteristics. The local ethics committee approved the study. Participants gave written informed consent and were reimbursed for participation.

Measures of addiction severity. Addiction severity was assessed using (1) time-line follow back score (TLFB; Sobell, 1992), to assess alcohol units consumed in the month before treatment; (2) obsessive-compulsive drinking scale (OCDS; Anton et al., 1995); (3) alcohol craving questionnaire (ACQ; Tiffany et al., 2000); and (4) alcohol use disorder identification test (AUDIT; Allen et al., 1997).

Neurocognitive measurements. Alcohol dependence is known to be linked with a number of cognitive deficits (Bates et al., 2002; Goldstein et al., 2004), which have recently been shown to be associated with impaired model-based decision-making (Sebold et al., 2014). Therefore, participants completed a battery of neurocognitive tests on the following domains: working memory (Digit Span; Wechsler, 1955); cognitive speed (Digit-Symbol-Substitution Test; Wechsler, 1955); reasoning (Matrices Test; Amthauer et al., 1999); verbal IQ (German vocabulary test, Schmidt and Metzler, 1992); visual attention (Reitan Trail Making A; Reitan, 1955); and complex attention (Reitan Trail Making B; Reitan, 1955). Results and group comparisons are summarized in Table 1. We computed a factor analysis (principle component analysis) to extract composite measures of neurocognitive functioning. Based on an eigenvector cutoff of >1, a factor analysis with an oblique rotation (direct oblimin) yielded a single factor solution, accounting for 59.61% of variance in the six test results obtained. The composite measure of neurocognitive functioning was subsequently used as a covariate in control analyses.

Decision-making task. Participants performed reward-based decision-making in a dynamic environment that requires flexible behavioral adaptation (Fig. 2*A*, illustration). In a total of 160 trials, participants decided between two cards, each showing a different geometric stimulus (maximum response time, 1.5 s). Importantly, the task incorporated a simple higher-order structure: reward probabilities associated with the two choice options were anticorrelated; whenever stimulus A was a good choice, stimulus B would be the worse choice, and vice versa. Even though the outcome for the alternative option is never shown, the agent can infer from the anticorrelation of the options what might have happened if he had taken the other stimulus (Fig. 1*A–C*). Reward contingencies remained stable for the first 55 trials (first, "prereversal," phase) and also for the last 35 trials (last, "postreversal," phase). During the second

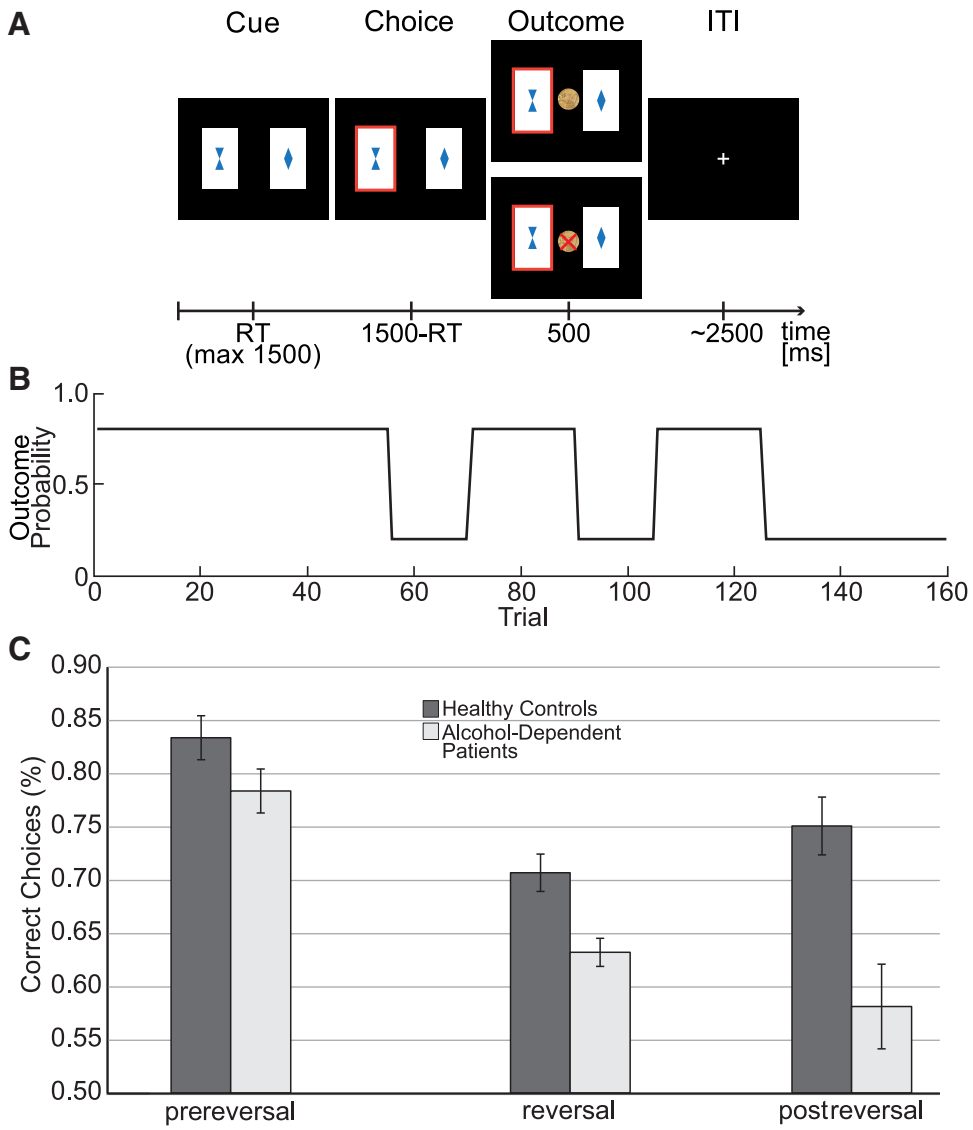

Figure 2. Decision-making task. ***A***, Exemplary trial sequence. ***B***, One of the stimuli was assigned with a reward probability of 80% and a punishment probability of 20% (vice versa for the other stimulus). Reward contingencies remained stable for the first 55 trials (prereversal block) and also for the last 35 trials (postreversal block). In between, reward contingencies changed four times (reversal block). ***C***, Raw data results. Correct choices differed significantly as a function of phase (prereversal, reversal, postreversal, $F = 21.78$, $p < 0.001$). We observed a main effect group and a significant interaction of phase × group ($F = 3.27$, $p = 0.04$). Between-group *post hoc t* tests revealed that group differences were present in the reversal phase ($t = 3.48$, $p = 0.001$) and in the postreversal phase ($t = 3.36$, $p = 0.001$), but not in the initial stable prereversal phase ($t = 1.69$, $p = 0.10$). Error bars indicate Standard Errors of the Mean.

("reversal") phase, reward contingencies changed (four changes in total, after 15 or 20 trials; Fig. 2*B*). This required participants to flexibly adapt their behavior.

Right-side versus left-side location of the stimuli on the screen was randomized over trials. After the participant had chosen one stimulus by left or right button press, the selected stimulus was highlighted and depicted for 1.5 s minus the reaction time. Feedback was shown for 0.5 s (monetary win vs monetary loss, indicated by a 10 Eurocent coin or a crossed 10 Eurocent coin, respectively). During the intertrial interval, a fixation cross was presented for a variable duration (jittered and exponentially distributed; range, 1–12.5 s). If no response occurred during the decision window, the message "too slow" was presented, and no outcome was delivered.

In a prior instruction and training session outside the MRI scanner, participants were informed that one of the two cards had a superior chance of winning money (probabilistic nature of the task). They were told that, depending on their choice, they could either win 10 cents or lose 10 cents per trial, that the aim was to win as much as possible, and that the total amount of money gained would be paid out at the end of the experiment. Participants performed 20 training trials with a different set of cards and without any reversal of reward contingencies. Subsequently, participants were instructed that reward probabilities could change over the course of the main experiment and that they should track such changes to win as much money as possible. Importantly, no other information or details on reversals or the correlation of outcomes was provided, such that patients had no explicitly instructed knowledge about the anticorrelated task structure before the experiment.

Analysis of choice behavior. Behavioral performance was quantified as the percentage of correct choices (choices of the stimulus with 80% reward probability) and was analyzed using repeated-measures ANOVA including the between-subject factor "group" (patients vs control subjects) and the within-subject factor "phase" (prereversal: first 35 trials; reversal: intermediate 90 trials; postreversal: last 35 trials).

We additionally investigated the effect of previous feedback on subsequent decisions, namely repeating choices after reward ("win–stay") and shifting responses after losses ("lose–shift"). Furthermore, we quantified how often participants repeated a choice despite two consecutive losses

for the same choice in the preceding two trials, relative to all loss trials (den Ouden et al., 2013).

Computational modeling. Different RL models were fitted to the data. All models learn the values of choice options via reward prediction errors (RPEs), a teaching signal that compares received rewards and expected values. In essence, the first three RL models differ in the degree of updating both the chosen and alternative decision options, as follows: (1) a model-free learner updating values for the chosen stimulus only, which neglects the anticorrelated task structure, which we refer to as the single-update (SU) model; (2) a learner updating values of chosen and unchosen stimuli equally using inference on the anticorrelated task structure, which we refer to as the double-update (DU) model; and (3) a model connecting SU and DU models by individually weighting the degree of double-update learning, thus accounting for individual variability. This is given by the weighting parameter κ. In the following, we refer to this model as the iDU model.

First, the model-free SU-algorithm updates a decision value $Q_{a,t}$ for the chosen stimulus via the RPE $\delta_{Q_{a,t}}$, which is defined as the difference between the received reward R_t and the anticipated reward for the chosen stimulus $Q_{a,t}$:

$$\delta_{Q_{a,t}} = R_t - Q_{a,t} \tag{1}$$

The RPE $\delta_{Q_{a,t}}$ is used to iteratively update decision values of the chosen decision value trial-by-trial:

$$Q_{a,t+1} = Q_{a,t} + \alpha\delta_{Q_{a,t}} \tag{2}$$

Here, α depicts the learning rate, which weights the influence of RPEs $\delta_{Q_{a,t}}$ on the updated values. α has natural boundaries between 0 and 1. Importantly, this model neglects the anticorrelated task structure by updating only decision values for the chosen stimulus, while the value of the alternative, unchosen stimulus $Q_{ua,t}$ remains unchanged, as follows:

$$Q_{ua,t+1} = Q_{ua,t} \tag{3}$$

Second, the DU algorithm updates chosen and unchosen decision values in each trial. This takes into account the anticorrelated structure of the task. In our modeling approach, this is captured by additionally updating the unchosen decision values based on a different error signal, which compares the fictive outcome that might have happened with the value of the unchosen option. The RPE for the DU model is as follows:

$$\delta_{Q_{ua,t}} = -R_t - Q_{ua,t} \tag{4}$$

The same learning rate α is used for updating unchosen values, as follows:

$$Q_{ua,t+1} = Q_{ua,t} + \alpha\delta_{Q_{ua,t}} \tag{5}$$

Equation 5 gives the same weight to the update of unchosen decision values as to the chosen decision values. Third, and in contrast, we assume that the degree of updating the alternative choice option differs across individuals. To account for interindividual variability regarding this process, we additionally constructed an iDU model to quantify each individual's degree of DU learning. This is provided by the parameter κ, which weights the learning rate α for the unchosen RPE $\delta_{Q_{ua,t}}$:

$$Q_{ua,t+1} = Q_{ua,t} + \kappa\alpha\delta_{Q_{ua,t}} \tag{6}$$

In the iDU model, the RPE $\delta_{Q_{a,t}}$ is weighted by the product of the learning rate for the chosen value and the weighting parameter κ, where $\kappa = 0$ reduces to the SU model, and $\kappa = 1$ to the DU model. Note that this results in lower learning rates for DU learning, which is in line with the key assumption that double-update learning is computationally more costly.

Figure 1 provides a schematic of inference on the anticorrelated task structure with respect to unchosen choice values (double-updating). In the task at hand, as double-updating depends on inference derived from actually experienced feedback, updating of the unchosen stimulus always relies on learning from feedback for the chosen stimulus (i.e., is rather unlikely to be a process independent from updating the chosen stimulus; for comparison with an identical implementation, see Li et al., 2011). We ran 1000 simulations of choices on the reward sequences of the empirical data via the used RL models by setting $\kappa = 0$, $\kappa = 0.5$, and $\kappa = 1$, and confirmed clear superiority of double updating in terms of correct choices in the middle reversal phase (68.60% correct choices for $\kappa = 0$; 75.50% for $\kappa = 0.50$; and 75.66% for $\kappa = 1$).

For tasks such as the one used here, some previous work indicated that models with a dynamically changing learning rate might fit behavior better than models with a static learning rate (Krugel et al., 2009). The so-called Sutton-K1 model updates the learning rate dynamically as a function of the change in prediction errors encountered (Sutton, 1992). It was previously discussed and used as a non-hierarchical approximation of a dynamic learning rate (Chumbley et al., 2012; Kepecs and Mainen, 2012; Landy et al., 2012; Iglesias et al., 2013). By including this model, we tested whether a model with a dynamic learning rate captures the observed behavior better than algorithms with a fixed learning rate. In this model, values are also updated via prediction errors as in Equations 1 and 2. The dynamic learning rate is transformed with a logistic function to remain in boundaries between 0 and 1, as follows:

$$\alpha_1 = \frac{1}{1 + \exp(-\iota_t)} \tag{7}$$

This is initialized with $\iota = 0$ corresponding to an initial learning rate of 0.5. Note that this parameter is called β in the original publication, which we here change to ι because β is used throughout the article to refer to the temperature in the decision model. The update of ι for the next trial depends on the change in reward prediction errors where:

$$\iota_{t+1} = \iota_t + \mu\delta_{Q_{a,t}} h_t, \tag{8}$$

and

$$h_{t+1} = (h_t + \alpha_t + \delta Q_{a,t}) * \max((1 - \alpha_t), 0). \tag{9}$$

The value of μ given in Equation 8 is a free parameter, which controls the individual degree of dynamic update of the learning rate. ι is a sensitivity parameter of the learning rate, controlling the influence of the RPE of the last trial on a trial-by-trial basis as a function of μ.

In sum, we had a total of four learning algorithms, namely SU, DU, iDU, and Sutton-K1. In all algorithms, we include the initial value of one option as a free parameter (Huys et al., 2011, 2012; Schlagenhauf et al., 2014).

Decision model. For all models, decisions are transformed into action probabilities by applying a softmax equation. The softmax equation includes the temperature β, which reflects the stochasticity of the choices; and a′ indicates all available choice options:

$$p(a) = \frac{\exp(\beta Q(a))}{\sum \exp(\beta Q(a'))} \tag{10}$$

Learning from rewards versus punishments. We also aimed to test the hypothesis of whether a potential deficit of alcohol-dependent patients in DU learning differs specifically as a function of learning from rewards versus learning from punishments in our task. In our models, we account for this by estimating separate learning rates and temperatures for reward (rew) and punishment (pun) trials corresponding to α_{rew}, α_{pun}, and β_{rew}, β_{pun}, respectively. We did so by assuming that these trial types refer to rather categorical differences in how tightly learned values influence choices.

Model fitting. Fitting was performed in the same Bayesian framework as introduced in the studies by Huys et al. (2011, 2012) and as used in several studies, including between-group designs (Chowdhury et al., 2013; Deserno et al., 2015a) and patient studies (Schlagenhauf et al., 2014; Deserno et al., 2015c). To infer the maximum a posteriori estimate of parameters θ for each individual i, we use a Gaussian prior with mean and variance μ and σ, as follows:

$$\mathrm{MAP}_i = \mathrm{argmax} \log p(Y \mid \theta)\, p(\theta \mid \mu, \sigma), \tag{11}$$

where Y represents the data in terms of actions A_i per subject i. We set priors empirically to the maximum-likelihood estimates of μ and σ given the data by all subjects included, as follows:

$$ML_i = \text{argmax} \log p(Y \mid \theta), \quad (12)$$

and achieve this by using expectation maximization. Constrained parameters were transformed to a logistic (α, κ) or exponential (β) distribution to enforce constraints and to render normally distributed parameter estimates. All modeling analyses were performed using Matlab 2010b. It should be noted that the empirical prior mainly serves to mildly regularize parameters at the population level. As this was performed based on the data of participants, this renders between-group parameters valid.

Model comparison. For all models, we approximate the model evidence by integrating out free parameters. This integral was approximated by sampling from the empirical prior distribution (Huys et al., 2011, 2012). Due to the hierarchical fitting procedure, which also fits prior means and variances (see Model fitting), such marginalized likelihoods can lead to overly optimistic results by biasing model selection toward more complex models. To obviate this problem, we used leave-one-out cross-validation by fitting the data without subject k and then marginalizing for subject k via sampling from the empirical prior distribution of sample $n - k$. Then, the marginal or integrated likelihood (the model evidence) based on leave-one-out cross-validation was subjected to a random-effects Bayesian model selection procedure (spm_BMS function contained in SPM8; Stephan et al., 2009) to compute expected posterior probabilities (PPs) and exceedance probabilities (XPs) for each model. XPs give the probability that PPs of a model differ from that of another model in the comparison set. Importantly, after running BMS initially across all participants, this was then performed separately for control subjects and patients.

Adequacy of the best-fitting model. In addition to relative model comparison, we assessed the quality of the best-fitting model as follows: (1) determining identifiability via the rank of the Jacobian matrix (Bamber and van Santen, 1985, 2000) and via assessing correlations between the inferred parameters; (2) measures of absolute model fit via calculating McFadden's pseudo-R^2 and assessing how many of each participant's choices can be explained by the model (corresponds to each individual's negative log-likelihood relative to the number of trials (Daw, 2009; Huys et al., 2011, 2012), which was tested for significance against chance level using a binomial test); (3) simulating choice data (100 simulations/participant) of the task based on the inferred parameters and running the same behavioral analysis on simulated choice data (using the median of the 100 simulations/subject), as for the empirical data; and (4) refitting the simulated choice data as a recovery analysis and determining the correlation between parameters inferred from empirical data with parameters inferred from simulated data.

MRI data acquisition. Functional imaging was performed using a 3 tesla Siemens Trio scanner to acquire gradient echo T2*-weighted echoplanar images with blood oxygenation level-dependent contrast. Covering the whole brain, 40 slices were acquired in oblique orientation at 20° to the anterior commissure–posterior commissure line and in ascending order, with 2.5 mm thickness, 3×3 mm^2 in-plane voxel resolution, 0.5 mm gap between slices, TR = 2.09 s, TE = 22 ms, and flip angle $\alpha = 90°$. Before functional scanning, a field distortion map was collected to account for individual homogeneity differences of the magnetic field. Additionally, T1-weighted anatomical images were acquired.

Preprocessing of fMRI data. For fMRI data analysis, we used SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Images were corrected for delay of slice time acquisition. Voxel-displacement maps were estimated based on acquired field maps. To correct for motion, all images were realigned, and additionally corrected for distortion and the interaction of distortion and motion. The images were spatially normalized to Montreal Neurological Institute (MNI) space using normalization parameters generated during the segmentation of the individual T1-weighted structural image (Ashburner and Friston, 2005); thereafter, all images were spatially smoothed with an isotropic Gaussian kernel (6 mm full-width at half-maximum).

Statistical analysis of functional MRI. The aim of the statistical analysis was to elucidate neural signatures of RPEs for chosen values as a function of SU versus DU learning and potential group differences. Based on each individual's set of parameters identified during model fitting (random-effects parameters), we computed regressors for the statistical analysis of fMRI data. Using the general linear model approach as implemented in SPM8, smoothed images were analyzed in an event-related manner. At the first level, onsets of feedback were entered into the model and convolved with the canonical hemodynamic response function and modulated parametrically by two trial-by-trial regressors from our modeling analysis, as follows: first, individual RPEs for chosen values were computed based on of the SU model with $\kappa = 0$ (RPE_{SU}). Second, a difference regressor was entered reflecting the difference of $RPE_{DU} - RPE_{SU.}$ To build this regressor, individual RPEs for chosen values were computed based on the DU model with $\kappa = 1$ (RPE_{DU}) and subtracted from the RPE_{SU} described above. This procedure accounts for collinearity between the regressors and reflects unique variance due to double-update computations beyond the single-update RPE (for the same analytic approach, please compare with Daw et al., 2011). The difference between RPE_{DU} and RPE_{SU} reflects the difference in chosen values from the DU and SU algorithms. In the iDU model, this difference is reflected in the estimate of κ (illustrated in Fig. 1*C*). Throughout the article, the second parametric modulator (the difference regressor) is referred to as RPE_{DU}. Missing trials were modeled separately. To account for possible confounds due to movement, we included the six realignment parameters, the first temporal derivative of the translational realignment parameters and a further regressor censoring scan-to-scan movement of >1 mm. At the second level, contrast images for RPE_{SU} and RPE_{DU} were taken to a random-effects analysis. A full-factorial ANOVA contained the type of RPEs (RPE_{SU}/RPE_{DU}) as the within-subject factor, and group as the between-subject factor.

Voxel-based morphometry. Each subject's anatomical T1-weighted image was segmented into different tissue classes using the unified segmentation approach implemented in SPM8 (Ashburner and Friston, 2005). Modulated images of gray matter density were smoothed using an isotropic Gaussian kernel (6 mm full-width at half-maximum) and subjected to a random-effects model. The volume of gray matter, white matter, and CSF tissue classes were summed to gain an individual estimate of total intracranial volume, which was entered as a covariate in between-group comparisons. As there is strong evidence for pronounced cortical gray matter density loss in alcohol-dependent individuals (Beck et al., 2012), we first tested for differences in gray matter density between the patient group and the control group. The patient group was characterized by significantly reduced gray matter density (FWE corrected for the whole brain, $p < 0.05$) predominantly in a large cluster covering the cingulate cortex (see Table 7). Second, to control for differences in gray matter density as a potential confound of our fMRI results, we extracted gray matter density from the following two regions of interest: (1) based on the fMRI analysis, the conjunction of both RPEs across the entire sample (thresholded at FWE corrected, $p < 0.05$); and (2) an anatomically predefined mask of combining frontal lobe and cingulate cortex (obtained from AAL templates, WFUPickAtlas Toolbox).

Results

Behavioral raw data analyses

Correct choices

An ANOVA revealed a significant effect of phase ($F = 21.76$, $p < 0.001$) and group ($F = 19.97$, $p < 0.001$), and a significant group × phase interaction ($F = 3.27$, $p = 0.04$, Fig. 2*C*).

Win–stay and lose–shift

We further explored patients' deficit in correct choices by analyzing how often participants repeated choices after reward, "win–stay," and shifted after losses, "lose–shift." A between-group difference was observed on win–stay ($t = 2.23$, $p = 0.03$) with patients showing less stay behavior after wins (control subjects: mean, 0.93; SD, 0.06; patients: mean, 0.87; SD, 0.14). There was no difference in lose–shift ($t = 0.25$, $p = 0.80$).

Repeating choices despite recurrent negative consequences

We found a significant between-group difference ($t = 2.63$, $p = 0.01$) in repetition behavior after two successive losses (control subjects: mean, 0.11; SD, 0.08; patients: mean, 0.18; SD, 0.14); patients reiterated disadvantageous choices more often, despite negative consequences in preceding trials.

Computational modeling of behavior

Computational modeling: model comparison

Using random-effects Bayesian model selection (BMS) (Stephan et al., 2009) across control subjects and patients, the iDU model with κ as a free parameter, and with separate learning rates and temperatures for reward and punishment trials (α_{rew}, α_{pun}, β_{rew}, β_{pun}) peaked out of 14 models ($XP_{iDU} = 0.71$, $PP_{iDU} = 0.27$). The overall superiority of separate learning rates and temperatures for reward and punishment trials was also confirmed when grouping the 14 models in four families ($1\beta1\alpha$, $2\beta1\alpha$, $2\alpha1\beta$, and $2\alpha2\beta$), with the first two families containing four models each (SU, DU, iDU, and Sutton-K1) and the latter two each consisting of three models (SU, DU, and iDU) because it is not straightforward to define the dynamic learning rate separately for reward and punishments ($XP_{1\alpha1\beta} = 0$, $PP_{1\alpha1\beta} = 0.07$; $XP_{2\beta1\alpha} = 0.01$, $PP_{2\beta1\alpha} = 0.24$; $XP_{2\alpha1\beta} = 0$, $PP_{2\alpha1\beta} = 0.04$; $XP_{2\alpha2\beta} = 0.99$, $XP_{2\alpha2\beta} = 0.66$).

Importantly, when running BMS for both groups separately across all 14 models, control subjects and patients differed regarding the model that explained their behavior relatively better (Fig. 3*A*, Table 2); control subjects were best explained by the iDU model that includes inference on the task structure given by the parameter κ, an individual weight of the degree of DU learning ($XP_{iDU} = 0.92$, $PP_{iDU} = 0.27$). Patients were relatively better explained by the model-free SU algorithm, which neglects an update of the alternative choice option ($XP_{SU} = 0.83$, $PP_{SU} = 0.27$). We verified that these group differences were not driven by a small subgroup of patients. Looking at individual relative model fit, 23 of 35 healthy control subjects were better explained by the iDU model than by the SU model, 25 of 43 patients were relatively better explained by the SU model than by the iDU model. Details regarding BMS can be found in Table 2, including log-likelihoods, log model evidence, and PPs and XPs for all 14 models separately for control subjects and patients. As requested by one of our reviewers, we fitted both groups separately, and this confirmed the model selection results as described above (control subjects: $XP_{iDU2\alpha2\beta} = 0.56$; $PP_{iDU2\alpha2\beta} = 0.20$; alcohol-dependent patients: $XP_{SU2\alpha2\beta} = 0.76$; $PP_{SU2\alpha2\beta} = 0.27$).

Computational modeling: identifiability, absolute model fit, and simulated choice data

All reported quality checks refer to the iDU model, which was best fitting across all participants. First, the rank of the Jacobian matrix equaled the number of free parameters in the model, indicating the identifiability of the model (Bamber and van Santen, 1985, 2000). Correlations between all parameters were acceptable ($r \leq 0.36$); only temperatures showed relatively strong correlations at $r = 0.67$ but were, however, consistently different for win and loss trials, in terms of being higher for wins than losses in all but two individuals. Second, adjusted McFadden's pseudo-R^2 ($R^2 = 0.60$) indicated reasonable absolute model fit. Only one healthy participant and six alcohol-dependent patients were not fit better than chance by any of the tested models. Notably, all relative model comparison results reported above were robust against excluding these participants who were not fitted better than chance (control subjects: $XP_{iDU2\alpha2\beta} = 0.92$, $PP_{iDU2\alpha2\beta} = 0.27$; patients: $XP_{SU2\alpha2\beta} = 0.79$, $PP_{SU2\alpha2\beta} = 0.28$). Third, choice data were simulated based on the inferred parameters of the best-fitting iDU model and tested in the same manner as the original empirical data to establish whether the model replicates group differences on choice behavior (correct choices, win–stay, repetition of punished actions). As we were interested in the replication of the empirically found effect, hypotheses were directed, and one-tailed tests were used. The model replicated the main effect of group on correct choices ($t = 3.52$, $p < 0.001$), as well as the group effect on win–stay rates ($t = 3.20$, $p < 0.001$) and on repetition behavior after punishment ($t = 1.86$, $p = 0.03$). Inferred model parameters did not recover the group × phase interaction observed in the raw data. Fourth, when refitting the simulated choice data, we observed strong correlations of modeling parameters derived from the empirical data and modeling parameters derived from the simulated data (correlation coefficients: $\alpha_{rew} = 0.55$, $\alpha_{pun} = 0.87$, $\beta_{rew} = 0.65$, $\beta_{pun} = 0.79$, $\kappa = 0.80$, $Q_i = 0.74$).

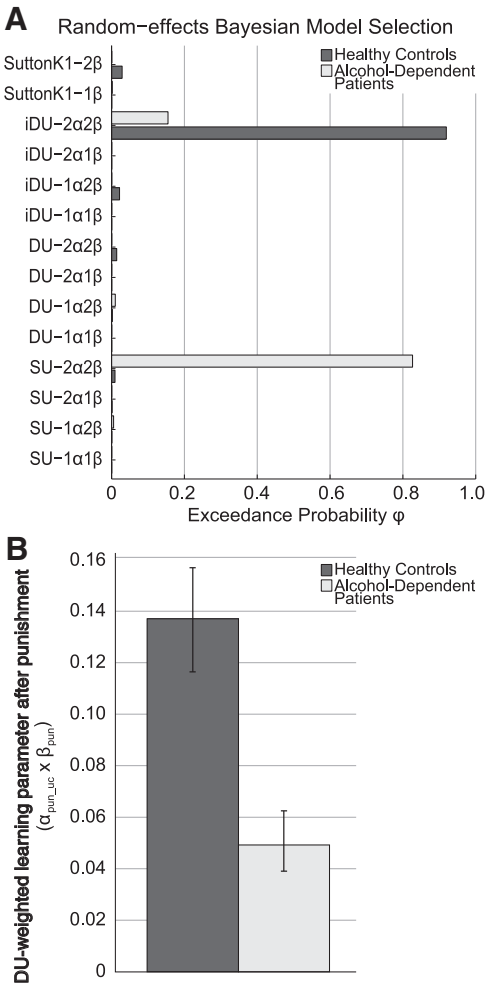

Figure 3. Computational modeling results. ***A***, Bayesian model selection revealed that healthy control subjects were best explained by the iDU model, including a factor which weights the individual degree of inference (double-updating), whereas for alcohol-dependent patients, model evidence was maximal in favor of the model-free single-update model. Models with separate learning rates and temperatures for reward and punishment trials outperformed models without this distinction. ***B***, Between-group comparisons on the inferred parameters derived from the best-fitting model (MANOVA) revealed a significant group difference on the parameters ($F = 2.83$, $p = 0.03$). *Post hoc* tests showed that the iDU punishment parameter was significantly lower in alcohol-dependent subjects compared with healthy control subjects ($F = 7.89$, $p = 0.006$). There were no significant group differences in any of the other inferred parameters of the model. Error bars indicate Standard Errors of the Mean.

Table 2. Model selection results

	1α1β		1α2β		2α1β		2α2β		Sutton1β		Sutton2β	
	HC	ALC	HC	ALC	HC	ALC	HC	ALC	HC	ALC	HC	ALC
SU												
LL	−2993.93	−1677.98	−2755.87	−1739.54	−2941.53	−1639.73	−2710.18	−1879.34	−3089.06	−1747.73	−2861.72	
ML	−3118.23	−1829.92	−2948.64	−1913.68	−3072.59	−1810.17	−2931.40	−1951.19	−3160.13	−1845.10	−3026.89	
PP	0.0193	0.0467	0.0914	0.0553	0.0214	0.0872	0.2678	0.0251	0.0730	0.1136	0.0366	
XP	0.0000	0.0008	0.0057	0.0015	0.0000	0.0093	0.8259	0.0001	0.0019	0.0294	0.0001	
DU												
LL	−2997.30	−1664.18	−2773.74	−1792.02	−2963.30	−1670.50	−2764.38					
ML	−3201.00	−1857.98	−3059.58	−1974.00	−3242.14	−1855.03	−3086.95					
PP	0.0204	0.0617	0.1035	0.0239	0.0191	0.0961	0.0485					
XP	0.0000	0.0022	0.0105	0.0001	0.0000	0.0143	0.0003					
iDU												
LL	−2875.53	−1562.88	−2628.62	−1668.27	−2848.17	−1546.36	−2584.38					
ML	−3056.65	−1771.15	−2935.57	−1870.86	−3052.73	−1756.96	−2888.28					
PP	0.0349	0.1063	0.0458	0.0340	0.0373	0.2685	0.1811					
XP	0.0001	0.0221	0.0002	0.0003	0.0001	0.9195	0.1552					

All models were compared using Bayesian model selection. We report log likelihoods (LLs), marginalized log likelihood (ML), exceedance probabilities (XPs), and expected posterior probabilities (PPs). HC, healthy control subjects; ALC, alcohol-dependent patients.

Table 3. iDU model: best fitting parameters

	β_{reward}	β_{punish}	α_{reward}	α_{punish}	$\kappa^*\alpha_{reward}$	$\kappa^*\alpha_{punish}$
Healthy control subjects	4.29 ± 1.18	2.04 ± 1.48	0.58 ± 0.22	0.47 ± 0.25	0.10 ± 0.11	0.09 ± 0.12
	25th P = 3.16	25th P = 1.00	25th = 0.43	25th = 0.29	25th P = 0.02	25th P = 0.01
	50th P = 4.55	50th P = 1.75	50th = 0.59	50th = 0.45	50th P = 0.06	50th P = 0.04
	75th = 5.02	75th P = 2.52	75th = 0.76	75th = 0.71	75th P = 0.13	75th P = 0.10
Alcohol-dependent patients	4.24 ± 0.98	1.58 ± 1.22	0.51 ± 0.31	0.47 ± 0.32	0.07 ± 0.11	0.04 ± 0.05
	25th P = 3.62	25th P = 0.84	25th P = 0.22	25th P = 0.11	25th P = 0.01	25th P = 0.01
	50th P = 4.19	50th P = 1.31	50th P = 0.49	50th P = 0.54	50th P = 0.02	50th P = 0.02
	75th P = 4.99	75th P = 1.68	75th P = 0.82	75th P = 0.72	75th P = 0.08	75th P = 0.06

Data are reported as the mean ± SD, unless otherwise indicated. Only multiplications of decision noise beta with learning parameters are reported in the Results section. P, Percentile.

Computational modeling: group differences on model parameters

We tested for between-group differences in individuals fit better than chance by subjecting the inferred parameters of the iDU model, the best-fitting model across both groups (Table 3), to a multivariate ANOVA (MANOVA) with group as the between-subject factor (patients vs control subjects). This MANOVA contained the following parameters, each separately for reward and punishment: learning rates for the update of chosen (α_{rew_c}, α_{pun_c}) and unchosen values (α_{rew_uc}, α_{pun_uc}, products of the weighting factor κ with α_{rew_c}, and α_{pun_c}), each multiplied by the temperature for reward or punishment trials, respectively. This revealed a significant effect of the between-subject factor group ($F = 2.83$, $p = 0.03$). We explored this group difference using *post hoc t* tests to compare each of the parameters between groups. In line with the raw data results, we found a significantly lower DU-weighted punishment parameter ($\alpha_{pun\_uc} \times \beta_{pun}$, $F = 7.89$, $p = 0.006$; Fig. 3*B*), whereas none of the other parameters differed significantly between groups (group differences regarding learning rates of the simpler model-free SU algorithm, all $p > 0.66$). Note that the group difference on the DU-weighted punishment parameter was also present when comparing parameters derived from the less well fitting model, with only one temperature parameter for both reward and punishment trials ($t = 2.35$, $p = 0.02$).

Association of modeling parameters with repetition of choices despite recurrent punishment

A multiple regression model with the perseveration score (repeating choices despite recurrent punishment) as the dependent variable, and the DU and SU parameters for reward and punishment, respectively, as predictors ($R^2 = 0.27$; adjusted $R^2 = 0.24$) revealed a significant negative association specifically of the DU punishment parameter with the perseveration score ($\beta = -0.41$, $t = 2.79$, $p = 0.002$). All other parameters did not significantly predict the perseveration score (all t values $<|1.69|$, all p values >0.10). This suggests that a deficit in double-update punishment learning, possibly conflated with decision noise in these very trials, as indicated by selective between-group differences in modeling parameters (κ by α_{pun_c} by β_{pun}) explains perseveration after recurrent punishment.

Functional imaging results

Neural signatures of single- and double-update learning

To explore neural signatures of this behavioral deficit, we analyzed the encoding of two types of RPEs for the chosen option, namely RPE_{SU} versus RPE_{DU}. Effects for both types of learning signatures and their conjunction across both groups are illustrated in Figure 4, and in Tables 4, 5, and 6. For between-group differences, we tested for a type of RPE (RPE_{SU}/RPE_{DU}) × group (patients/control subjects) interaction. The conjunction of both RPEs across the entire sample (thresholded at FWE-corrected $p < 0.05$ for the whole brain; Fig. 4, Table 6) was used to correct for multiple comparisons (at FWE-corrected $p < 0.05$ based on this search volume). The RPE type × group interaction reached significance in the mPFC ($X = -10$, $Y = 62$, $Z = 12$; $t = 3.98$; FWE-corrected for the conjunction, $p = 0.01$) and posterior cingulate cortex ($X = 0$, $Y = -40$, $Z = 32$; $t = 3.72$; FWE-corrected for the conjunction, $p = 0.03$). As *post hoc* contrast, we compared RPE_{SU} and RPE_{DU} between groups. This confirmed significantly reduced

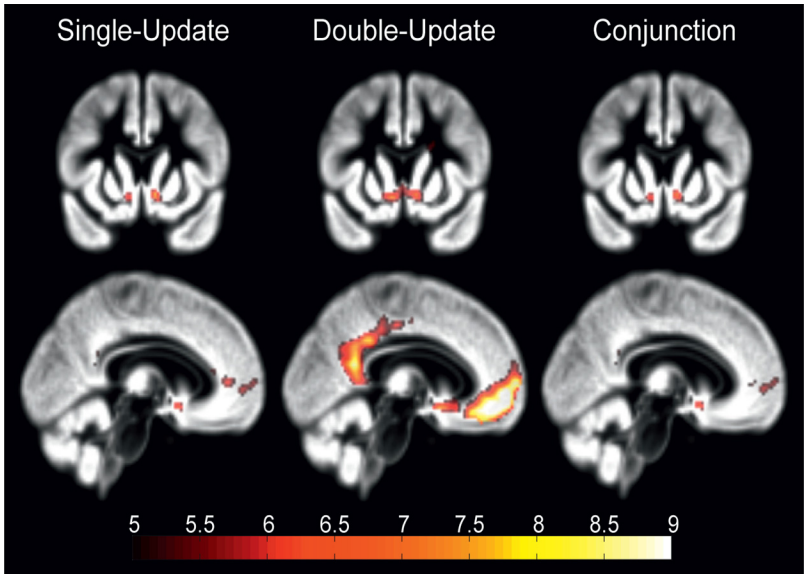

Figure 4. Neural coding of single-update vs double-update signals across the entire sample. Across all participants (patients and control subjects), we observed model-free RPE_{SU} in bilateral ventral striatum, and medial and lateral prefrontal cortex (FWE-corrected for the whole brain, $p < 0.05$; Table 4). For the difference regressor RPE_{DU}, we found effects in overlapping regions (bilateral ventral striatum, medial and lateral prefrontal cortex) and additionally in hippocampus and insula (FWE corrected for the whole brain, $p < 0.05$; Table 5). The conjunction of both contrasts revealed overlapping effects of RPE_{SU} and RPE_{DU}, in bilateral ventral striatum, medial and lateral prefrontal cortex, and posterior cingulate cortex (FWE corrected for the whole brain, $p < 0.05$; Table 6). The latter was used as a search volume for small-volume correction of group differences. Effects are reported using a significance level of $p < 0.05$, FWE corrected for the whole brain. Activations are shown superimposed on an averaged gray matter mask of the entire sample. For display purposes, threshold is set at $t > 5$.

Table 4. Neural signatures of single-update learning (RPE_{SU}) for both healthy control subjects and alcohol-dependent patients taken together at $p < 0.05$ FWE whole brain corrected

	Single-update signals			
	MNI coordinates	Cluster size	*t*	Peak *p* value (FWE corrected)
Ventral striatum	−8, 8, −10	57	8.66	<0.001
Ventral striatum	12, 8, −10	82	8.54	<0.001
Middle orbital gyrus	6, 42, −8	201	7.89	<0.001
Middle orbital gyrus	8, 60, 4		6.18	<0.001
Superior medial gyrus	−10, 64, 12	80	6.00	0.001
Middle orbital gyrus	−6, 54, 2		5.85	0.002
Anterior cingulate gyrus	−6, 44, 6	34	6.10	0.001
Anterior cingulate gyrus	−4, 30, 16	14	5.62	0.004
Middle orbitofrontal gyrus	−24, 32, −16	20	5.55	0.006
Putamen	−26, −6, 6	21	5.62	0.004
Putamen	26, 0, 2	17	5.53	0.006
Posterior cingulate gyrus	0, −34, 34	68	6.30	<0.001
Precuneus	−4, −50, 16	29	5.55	0.006
Angular gyrus	−46, −70, 34	17	5.46	0.008
Cerebellum	−44, −74, −34	161	6.80	<0.001
Cerebellum	36, −72, −40	93	6.20	<0.001
Cerebellum	44, −72, −32		5.49	0.008

coding of RPE_{DU} signatures in patients in the mPFC ($X = -8$, $Y = 62$, $Z = 12$; $t = 4.36$; FWE-corrected for the conjunction, $p = 0.003$; X = −6, $Y = 56$, $Z = 12$; $t = 3.68$; FWE-corrected for the conjunction, $p = 0.02$; Fig. 5) and posterior cingulate cortex ($X = -2$, $Y = -42$, $Z = 32$, $t = 3.72$; FWE-corrected for the conjunction, $p = 0.03$) but no significant between-group differences in activation elicited by model-free RPE_{SU}. We verified that the result of significantly reduced coding of RPE_{DU} signatures in patients in the mPFC was robust against excluding participants that were not fitted better than chance by any of the models. Indeed, when excluding these $n = 7$ participants, the group difference remained significant ($X = -8$, $Y = 62$, $Z = 12$; $t = 4.24$; FWE-corrected for the conjunction, $p_{peak} = 0.001$; and $X = -6$, $Y = 56$, $Z = 12$; $t = 3.78$; FWE-corrected for the conjunction, $p_{peak} = 0.011$).

In further analyses, we were interested in exploring associations of the observed reduced neural representation of RPE_{DU} in the mPFC with observed behavioral deficits and symptoms. Thus, mean parameter estimates at the peak of the between-group difference ($X = -8$, $Y = 62$, $Z = 12$, surrounded with an 8 mm sphere) were extracted to correlate them, for both groups separately, with the DU punishment parameter α_{pun_uc} by $\beta_{\_pun}$. Note that this approach is valid as we were specifically interested in associations of the neural reduction observed in patients versus control subjects (i.e., the neural group difference) with patients' behavioral deficit and symptoms. We did, however, not use the peak coordinate of the group difference to test further between-group hypotheses on the neural level, which would lead to circular inference, or "double dipping" (Kriegeskorte et al., 2009).

In patients, this revealed a positive association indicating that the attenuated mPFC double-update learning signature was related to a lower DU punishment parameter (Spearman's $r = 0.493$, $p = 0.006$; Fig. 5*C* left panel). No significant correlation was found in control subjects (Spearman's $r = 0.090$, $p = 0.61$). This confirms a link between the observed behavioral deficit in updating alternative options after punishment and the reduction of DU signatures in mPFC in patients.

Relationship between mPFC double-update learning signatures and symptom severity

We tested for an association of the reported neural alterations with symptom severity in alcohol-dependent patients. We performed a linear regression analysis with mean parameter estimates of the global maximum of the observed group difference in mPFC (at the peak voxel $X = -8$, $Y = 62$, $Z = 12$, with an 8-mm-radius sphere) as the dependent variable and the applied self-rating measurements of addiction severity (Table 1) as predictor variables, as follows: (1) units of alcohol consumed within 4 weeks before treatment commenced (TLFB); (2) OCDS; (3) ACQ; and (4) AUDIT. This revealed the OCDS score as having a significant negative association with the neural mPFC DU learning signature ($\beta = -0.64$, $t = 2.64$, $p = 0.01$; Fig. 5*C* right panel). Patients reporting a higher level of obsessive-compulsive drinking habits showed, on the neural level, lower coding of inference components regarding unchosen choice options. An additional regression model with the same independent variables and the DU punishment parameter as the dependent variable did not indicate any significant results (p values >0.52).

Table 5. Neural signatures of double-update learning (RPE_{DU}) for both healthy control subjects and alcohol-dependent patients taken together at $p < 0.05$ FWE whole brain corrected

	Double-update signals			
	MNI coordinates	Cluster size	*t*	Peak *p* value (FWE corrected)
Middle orbital gyrus	−2, 56, −4	2681	11.20	<0.001
Rectal gyrus	−6, 44, −10		10.52	<0.001
Inferior frontal gyrus	−34, 36, −10		10.20	<0.001
Inferior frontal gyrus	34, 36, −12	86	6.77	<0.001
Superior frontal gyrus	−12, 48, 36	190	6.65	<0.001
Superior frontal gyrus	−18, 38, 50		5.79	0.002
Middle frontal gyrus	−24, 32, 50		5.76	0.002
Insula	−38, −2, 14	37	5.88	0.001
Ventral striatum	−6, 8, −10	386	7.43	<0.001
Ventral striatum	10, 12, −8		7.01	<0.001
Anterior cingulate cortex	2, 20, 0		6.66	<0.001
Caudate	20, 18, 26	52	5.59	0.005
Hippocampus	−30, −12, −18	188	7.61	<0.001
Hippocampus	32, −28, −10		5.88	0.001
Fusiform gyrus	−32, −36, −14		5.5	0.007
Hippocampus	38, −24, −14	52	6.44	<0.001
Fusiform area	42, −18, −18		5.88	0.001
Posterior cingulate gyrus	−2, −42, 32	1201	9.05	<0.001
Precuneus	−6, −54, 18		8.7	<0.001
Posterior cingulate gyrus	−4, −52, 26		8.14	<0.001
Middle temporal gyrus	−50, −70, 22	180	6.23	<0.001
Angular gyrus	−46, −72, 32		5.96	0.001
Middle temporal gyrus	−44, −60, 22		5.92	0.001
Middle temporal gyrus	58, −8, −22	29	6.25	<0.001
Middle temporal gyrus	−60, −10, −20	52	6.12	0.001
Superior temporal gyrus	68, −22, 14	17	5.5	0.016
Superior temporal gyrus	60, −24, 16		5.19	0.025
Temporal pole	56, 2, 6	22	5.79	0.002
Operculum	44, −20, 20	40	5.67	0.004
Precentral gyrus	42, −16, 62		6.14	<0.001
Postcentral gyrus	40, −26, 58	69	5.07	0.038

Table 6. fMRI whole-brain results for the conjunction of single-update and double-update learning signals across both groups

Region	MNI coordinates	*k*	Cluster *p* value (FWE corrected)	*t* value	Peak *p* value (FWE corrected)
Superior medial gyrus	−10, 64, 12			6	0.001
Middle orbital gyrus	−6, 54, 2	79	<0.001	5.85	0.002
Middle orbital gyrus	6, 42, −8			7.46	<0.001
Middle orbital gyrus	8, 60, 4	174	<0.001	6.18	<0.001
Middle orbital gyrus	−24, 32, −16	19	0.001	5.55	0.006
Ventral striatum	−8, 8, −10	43	<0.001	7.39	<0.001
Ventral striatum	10, 12, −8	66	<0.001	7.01	<0.001
Posterior cingulate gyrus	0, −34, 34	56	<0.001	6.30	<0.001
Precuneus	−4, −50, 16	25	<0.001	5.55	0.006

Only clusters $k > 15$ are depicted.

Covariance analyses for possible confounding factors

To adjust for possible confounding influences, the following variables were included as covariates in the behavioral (correct choices and the DU punishment learning parameter) and fMRI analyses (RPE type × group interaction, group difference on RPE_{DU} coding): smoking status; depression score (Beck's depression inventory; Beck et al, 1996); and the composite measure of neurocognitive functioning as well as gray matter density (voxel-based morphometry, based on a functional and an anatomically predefined mask of frontolimbic structures). All reported results remained significant when adjusting for these possible confounds (all *p* values <0.05).

Table 7. Voxel-based morphometry: group differences

	Voxel-based morphometry			
	MNI coordinates	Cluster size	*t*	Peak *p* value (FWE corrected)
Supplementary motor cortex	2, 4, 48	523	7.59	<0.001
Superior medial gyrus	2, 30, 38		7.587.11	<0.001
Middle cingulate cortex	0, 12, 38			<0.001
Superior medial gyrus	4, 52, 38	26	7.44	<0.001
Middle cingulate cortex	0, −32, 46	378	7.25	<0.001
Middle cingulate cortex	2, −40, 50		6.53	0.001
Precuneus	0, −46, 38		6.44	0.002
Precuneus	2, −78, 40	22	6.41	0.002
Anterior cingulate	−2, 48, 16	24	6.36	0.003
Frontal pole	50, 48, 26	24	5.93	0.012

Control subjects > alcohol-dependent patients at FWE whole brain corrected $p < 0.05$.

Discussion

We provide novel insight into mechanisms of maladaptive decision-making and behavioral adaptation in patients with alcohol dependence and its underlying neural substrates. Our results support the view of intact model-free learning and behavioral control in addiction associated with a deficit in using environmental structure to guide decision-making: choice behavior in patients was best explained by a model-free RL algorithm, which neglects the updating of alternative choice options. This was due to a specific reduction in the updating of the alternative option after punishments. On the neural level, the learning signature of such a double-updating mechanism was reduced in patients' mPFC and correlated with the observed behavioral deficit in updating alternative choices as well as obsessive-compulsive drinking habits.

Disrupted behavioral adaptation in addiction

Deficits in cognitive flexibility are known in patients experiencing addiction (Bechara and Damasio, 2002; Garavan and Stout, 2005; Ersche et al., 2011; Goldstein and Volkow, 2011). In line with this, we demonstrate that alcohol-dependent patients show diminished behavioral adaptation in a dynamic environment. Crucially, by using computational modeling, we provide a mechanistic account for this deficit: alcohol-dependent patients are specifically impaired in their capacity to integrate alternative choice options and to accurately track the value of an alternative option after having received punishment. Put differently, patients show less consideration of "what might have been good instead": formally, after patients had received punishment for the chosen option, they did not increase the values of the alternative option as would have been appropriate according to the anticorrelated task structure, which was captured by a significantly lower double-update punishment parameter in patients. This finding derived from computational modeling can account for the overall impairment in correct decisions, reduced win-staying, and the repetition of choices despite successive punishment, as suggested by our simulation analysis. Therefore, our observation suggests that simpler, model-free, single-update learning is intact in addiction (such that the updating of chosen values after rewards and punishments remains relatively unaffected), but that updating of alternative, unchosen values is abolished after punishment. Such inference on what might have happened goes awry when values need adjustment after negative feedback, and thus potentially advantageous alternative choice options are neglected when making decisions. The finding is in line with recent animal models of addiction suggesting a specific deficit in mentally simulating outcomes not directly experienced and a disturbed integration of multiple predictions (Lucantonio et al., 2014).

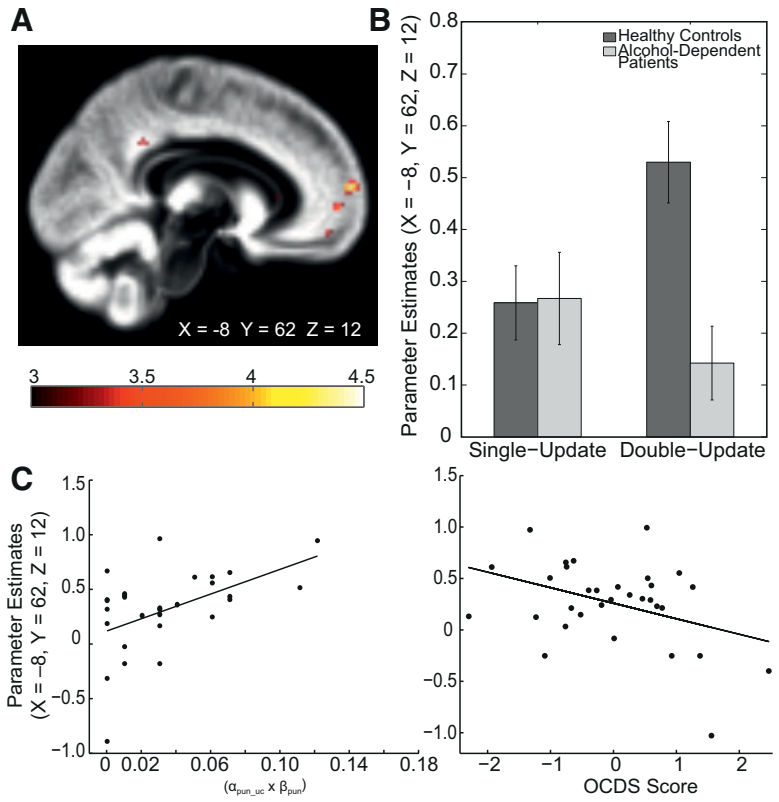

Figure 5. Group differences in the neural coding of single-update vs double-update signals. ***A***, Reduced inference signatures were found in the mPFC in alcohol-dependent patients compared with healthy control subjects ($X = -8$, $Y = 62$, $Z = 12$; $t = 4.36$; FWE-corrected for the conjunction, $p = 0.003$; $X = -6$, $Y = 56$, $Z = 12$; $t = 3.68$; FWE corrected for the conjunction, $p = 0.02$) and posterior cingulate cortex ($X = -2$, $Y = -42$, $Z = 32$; FWE corrected for the conjunction, $p = 0.03$; $t = 3.72$). No group difference regarding model-free signatures was found. For display purposes, thresholded at $t > 3$. ***B***, Plot of parameter estimates at the peaks of the group difference in the mPFC. ***C***, In patients, parameter estimates from an 8-mm-radius sphere around the peak coordinate ($X = -8$, $Y = 62$, $Z = 12$) of the group difference correlated with the behavioral deficit in double-update learning after punishments (left: Spearman's $r = 0.49$, $p = 0.006$). A multiple regression model including all applied measures of disease severity as explanatory variables predicting these parameter estimates revealed the sum score of the obsessive-compulsive drinking scale as the only significant predictor (OCDS; $\beta = -0.64$, $t = 2.64$, $p = 0.014$; right). Here, we plot Studentized residuals of the OCDS with respect to other disease severity measures.

Intriguingly, this behavioral deficit resonates well with clinical observations and diagnostic criteria of addiction describing the maintenance of disadvantageous behaviors despite negative consequences. Importantly, our finding goes beyond previous studies on behavioral adaptation linking addiction to blunted neural responses associated with performance errors and reduced error awareness (Paulus et al., 2008; Goldstein et al., 2009): a disturbed inference process regarding the update of alternative choice options may constitute one plausible explanation for these deficits.

In studies applying tasks similar to the one used here, inference about alternative choice options has been previously linked to a goal-directed or model-based control system (Hampton et al., 2006; Bromberg-Martin et al., 2010). An alternative explanation includes that double-update inference does not arise from a full model-based system but rather reflects temporal difference learning about the relationship of choice values (Shohamy and Wagner, 2008; Wimmer et al., 2012; Doll et al., 2015). In this framework, our results could be interpreted as an impairment in generalizing from one stimulus to another. Either way, the capacity to simultaneously update multiple decision values, including those of unobserved outcomes, might be regarded as sine qua non for building and using an internal model of the environment, which is important for goal-directed or model-based control. Using sequential decision-making, reduced model-based behavioral control was observed in alcohol-dependent patients (Sebold et al., 2014), although this impairment was attenuated when adjusting for cognitive deficits. In the present study, the impairment in updating alternative choice options remained significant when adjusting for cognitive capacities, suggesting a specific characteristic for alcohol dependence rather than an epiphenomenon of a global impairment. Thus, our finding of reduced inferential capacities appends prominent theories proposing a shift from goal-directed to habitual behavioral control in addiction (Everitt and Robbins, 2005; Dayan, 2009; Lucantonio et al., 2012).

Blunted mPFC double-update learning signatures in alcohol-dependent patients

Patients were characterized by reduced coding of double-update RPE signals in mPFC. Reduced representation of these inference signatures in patients' mPFC was related to the observed behavioral deficit and to obsessive-compulsive drinking habits. In line with our findings, alcohol-dependent patients showed hypoactivation in a similar region for a contrast assessing goal-directed learning during a different instrumental learning task (Sjoerds et al., 2013). In healthy individuals, the medial prefrontal and orbitofrontal cortex is known to encode model-based values computed "on the fly," which allows behavioral flexibility (Haber and Behrens, 2014). In consonance with this, the mPFC has been identified as a key region for flexible behavioral adaptation and model-based evaluation (Hampton et al., 2006; Daw et al., 2011). Specifically, this region has been linked to the integration of computations from habitual and goal-directed systems (Lee et al., 2014). Interestingly, Lee et al. (2014) identified computational signals for the reliability of both systems in the mPFC. Reliability signals are thought to be used by an arbitration mechanism to allocate the degree of control exerted by one of the two systems at a given point in time. Our observation of reduced double-update signatures at nearby coordinates may support a view on behavioral control in addiction that Lee et al. (2014) invite in their discussion: a failure of the arbitration process, namely the ability to appropriately parse behavioral control between different modes. Remarkably, reduced coding of double-update inference components in alcohol-dependent patients' mPFC remained significant when adjusting for reductions in gray matter density supporting the view of a specific neural signature of abolished inference. This interpretation is strengthened by correlations of mPFC signatures with reduced double-update learning rates after punishment and obsessive drinking habits in patients. Together, reduced double-update prediction error coding in alcohol-dependent patients' mPFC may indeed account for their decreased behavioral flexibility and constitute

one piece in the puzzle of obsessive alcohol consumption despite negative consequences.

Neurochemical considerations

Blunted presynaptic dopamine function was found in alcohol-addicted patients (Martinez et al., 2005), and lower levels of ventral striatal presynaptic dopamine were demonstrated to be associated with a lower degree of model-based behavioral control and diminished coding of model-based prefrontal signatures during sequential decision-making (Deserno et al., 2015b). Thus, low levels of presynaptic dopamine could hypothetically explain the reported findings to some extent. Further, reduced dopamine D_2 receptor availability is among the best-established findings in addiction (Volkow et al., 1990; Heinz et al., 2004). Low levels of D_2 receptors were linked to an impairment of re-evaluating decisions via the prefrontal cortex after negative feedback (Frank et al., 2004; Goto and Grace, 2005). Recent evidence from an animal model indicates that chronic alcohol-induced malfunction of, specifically, mPFC D_2/D_4 receptors disrupts flexible behavioral adaptation (Trantham-Davidson et al., 2014), which is in consonance with the presented findings. Interestingly, a behavioral study in humans showed that genetic variability in dopaminergic neurotransmission relates to perseveration during reversal learning (den Ouden et al., 2013), also supporting the view that dopamine could at least partially account for the behavior observed in alcohol-dependent patients.

Limitations

Whether diminished inference about alternative choice options arises as a consequence of long-term alcohol consumption or reflects a predisposition factor for the development of addictive behavior cannot be elucidated by a cross-sectional design. Groups differed in terms of general cognition, smoking status, and gray matter density even though our results were robust when adjusting for these variables. Cross-sectional studies in at-risk populations (Ersche et al., 2010; Reiter et al., 2016), and longitudinal designs are warranted to track the influence of dysfunctional behavioral control systems across different stages in the development of addiction. It is to be noted that our model was not able to capture one specific aspect of the observed choice behavior, namely the group × phase interaction on correct choices due to particularly reduced performance in the middle and last phase. Additional analyses of reaction times, missed choices, and self-report data consistently showed that this was not due to a general decline in performance over the course of the experiment in patients. Apart from this aspect, all empirical choice data effects could be replicated by the model, and measures of absolute model fit and identifiability indicated that the applied models served as a good explanation for the observed behavior.

Although the best-fitting model was technically invertible, as indicated by our identifiability checks, we have to caution that there is a lack of specificity within these parameters with respect to which specific parameter determines certain aspects of the choice behavior, especially with regard to the decision noise and the learning rates. We therefore used multiplications of those parameters to ensure stable between-group comparisons (Daw, 2009)

Summary

In conclusion, after punishment, alcohol-dependent patients showed a deficit to infer and integrate alternative choice options in their decisions. Our data provide the first neuroimaging support for reduced coding of this double-update inference process in the mPFC—a key region for flexible behavioral control—underlying this deficit. The same mPFC signatures were negatively related to obsessive-compulsive drinking habits. The computational psychiatry account applied here improves our understanding of the perplexing question of why addicted individuals continue drug consumption despite negative consequences.

References

Allen JP, Litten RZ, Fertig JB, Babor T (1997) A review of research on the Alcohol Use Disorders Identification Test (AUDIT). Alcohol Clin Exp Res 21:613–619. CrossRef Medline

Amthauer RB, Liepmann D, Beauducel A (1999) Intelligenz-Struktur-Test 2000. Göttingen, Germany: Horgrefe.

Anton RF, Moak DH, Latham P (1995) The obsessive compulsive drinking scale: a self-rated instrument for the quantification of thoughts about alcohol and drinking behavior. Alcohol Clin Exp Res 19:92–99. CrossRef Medline

Ashburner J, Friston KJ (2005) Unified segmentation. Neuroimage 26:839–851. CrossRef Medline

Bamber D, van Santen JP (1985) How many parameters can a model have and still be testable? J Math Psychol 29:443–473. CrossRef

Bamber D, van Santen JP (2000) How to assess a model's testability and identifiability. J Math Psychol 44:20–40. CrossRef Medline

Bates ME, Bowden SC, Barry D (2002) Neurocognitive impairment associated with alcohol use disorders: implications for treatment. Exp Clin Psychopharmacol 10:193–212. CrossRef Medline

Bechara A, Damasio H (2002) Decision-making and addiction (part I): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. Neuropsychologia 40:1675–1689. CrossRef Medline

Beck A, Wüstenberg T, Genauck A, Wrase J, Schlagenhauf F, Smolka MN, Mann K, Heinz A (2012) Effect of brain structure, brain function, and brain connectivity on relapse in alcohol-dependent patients. Arch Gen Psychiatry 69:842–852. CrossRef Medline

Beck AT, Steer RA, Brown GK (1996) Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation.

Bromberg-Martin ES, Matsumoto M, Hong S, Hikosaka O (2010) A pallidus-habenula-dopamine pathway signals inferred stimulus values. J Neurophysiol 104:1068–1076. CrossRef Medline

Chiu PH, Lohrenz TM, Montague PR (2008) Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. Nat Neurosci 11:514–520. CrossRef Medline

Chowdhury R, Guitart-Masip M, Lambert C, Dayan P, Huys Q, Düzel E, Dolan RJ (2013) Dopamine restores reward prediction errors in old age. Nat Neurosci 16:648–653. CrossRef Medline

Chumbley JR, Flandin G, Bach DR, Daunizeau J, Fehr E, Dolan RJ, Friston KJ (2012) Learning and generalization under ambiguity: an fMRI study. PLoS Comput Biol 8:e1002346. CrossRef Medline

Daw ND (2009) Trial-by-trial data analysis using computational models. In: Affect, learning and decision making, attention and performance XXIII (Phelps EA, Robbins TW, Delgado M, eds), pp 3–38. New York: Oxford UP.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215. CrossRef Medline

Dayan P (2009) Dopamine, reinforcement learning, and addiction. Pharmacopsychiatry 42 [Suppl 1]:S56–S65. CrossRef Medline

den Ouden HE, Daw ND, Fernandez G, Elshout JA, Rijpkema M, Hoogman M, Franke B, Cools R (2013) Dissociable effects of dopamine and serotonin on reversal learning. Neuron 80:1090–1100. CrossRef Medline

Deserno L, Wilbertz T, Reiter A, Horstmann A, Neumann J, Villringer A, Heinze HJ, Schlagenhauf F (2015a) Lateral prefrontal model-based signals are reduced in healthy individuals with high trait impulsivity. Transl Psychiatry 5:e659. CrossRef Medline

Deserno L, Huys QJ, Boehme R, Buchert R, Heinze HJ, Grace AA, Dolan RJ,

Heinz A, Schlagenhauf F (2015b) Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proc Natl Acad Sci U S A 112:1595–1600. CrossRef Medline

Deserno L, Beck A, Huys QJ, Lorenz RC, Buchert R, Buchholz HG, Plotkin M, Kumakara Y, Cumming P, Heinze HJ, Grace AA, Rapp MA, Schlagenhauf F, Heinz A (2015c) Chronic alcohol intake abolishes the relationship between dopamine synthesis capacity and learning signals in the ventral striatum. Eur J Neurosci 41:477–486. CrossRef Medline

Doll BB, Simon DA, Daw ND (2012) The ubiquity of model-based reinforcement learning. Curr Opin Neurobiol 22:1075–1081.

Ersche KD, Turton AJ, Pradhan S, Bullmore ET, Robbins TW (2010) Drug addiction endophenotypes: impulsive versus sensation-seeking personality traits. Biol Psychiatry 68:770–773. CrossRef Medline

Ersche KD, Roiser JP, Abbott S, Craig KJ, Müller U, Suckling J, Ooi C, Shabbir SS, Clark L, Sahakian BJ, Fineberg NA, Merlo-Pich EV, Robbins TW, Bullmore ET (2011) Response perseveration in stimulant dependence is associated with striatal dysfunction and can be ameliorated by a D(2/3) receptor agonist. Biol Psychiatry 70:754–762. CrossRef Medline

Everitt BJ, Robbins TW (2005) Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 8:1481–1489. CrossRef Medline

First MB, Spitzer RL, Gibbon M, Williams J (2001) Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition with psychotic screen (SCID-I/P W/ PSY SCREEN). New York: New York State Psychiatric Institute

Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943. CrossRef Medline

Garavan H, Stout JC (2005) Neurocognitive insights into substance abuse. Trends Cogn Sci 9:195–201. CrossRef Medline

Gläscher J, Hampton AN, O'Doherty JP (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex 19:483–495. CrossRef Medline

Goldstein RZ, Volkow ND (2011) Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. Nat Rev Neurosci 12:652–669. CrossRef Medline

Goldstein RZ, Leskovjan AC, Hoff AL, Hitzemann R, Bashan F, Khalsa SS, Wang GJ, Fowler JS, Volkow ND (2004) Severity of neuropsychological impairment in cocaine and alcohol addiction: association with metabolism in the prefrontal cortex. Neuropsychologia 42:1447–1458. CrossRef Medline

Goldstein RZ, Craig AD, Bechara A, Garavan H, Childress AR, Paulus MP, Volkow ND (2009) The neurocircuitry of impaired insight in drug addiction. Trends Cogn Sci 13:372–380. CrossRef Medline

Goto Y, Grace AA (2005) Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. Nat Neurosci 8:805–812. CrossRef Medline

Haber SN, Behrens TE (2014) The neural network underlying incentive-based learning: implications for interpreting circuit disruptions in psychiatric disorders. Neuron 83:1019–1039. CrossRef Medline

Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci 26:8360–8367. CrossRef Medline

Heinz A, Siessmeier T, Wrase J, Hermann D, Klein S, Grüsser SM, Grüsser-Sinopoli SM, Flor H, Braus DF, Buchholz HG, Gründer G, Schreckenberger M, Smolka MN, Rösch F, Mann K, Bartenstein P (2004) Correlation between dopamine D(2) receptors in the ventral striatum and central processing of alcohol cues and craving. Am J Psychiatry 161:1783–1789. CrossRef Medline

Huys QJ, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. PLoS Comput Biol 7:e1002028. CrossRef Medline

Huys QJ, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP (2012) Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS Comput Biol 8:e1002410. CrossRef Medline

Huys QJM, Guitart-Masip M, Dolan R, Dayan P (2015) Decision-theoretic psychiatry. Clin Psychol Sci 3:400–421. CrossRef

Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, Stephan KE (2013) Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 80:519–530. CrossRef Medline

Izquierdo A, Jentsch JD (2012) Reversal learning as a measure of impulsive and compulsive behavior in addictions. Psychopharmacology (Berl) 219:607–620. CrossRef Medline

Kepecs A, Mainen ZF (2012) A computational framework for the study of confidence in humans and animals. Philos Trans R Soc Lond B Biol Sci 367:1322–1337. CrossRef Medline

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535–540. CrossRef Medline

Krugel LK, Biele G, Mohr PN, Li SC, Heekeren HR (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proc Natl Acad Sci U S A 106:17951–17956. CrossRef Medline

Landy MS, Trommershäuser J, Daw ND (2012) Dynamic estimation of task-relevant variance in movement under risk. J Neurosci 32:12702–12711. CrossRef Medline

Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. Neuron 81:687–699. CrossRef Medline

Li J, Daw ND (2011) Signals in human striatum are appropriate for policy update rather than value prediction. J Neurosci 31:5504–5511. CrossRef Medline

Lucantonio F, Stalnaker TA, Shaham Y, Niv Y, Schoenbaum G (2012) The impact of orbitofrontal dysfunction on cocaine addiction. Nat Neurosci 15:358–366. CrossRef Medline

Lucantonio F, Takahashi YK, Hoffman AF, Chang CY, Bali-Chaudhary S, Shaham Y, Lupica CR, Schoenbaum G (2014) Orbitofrontal activation restores insight lost after cocaine use. Nat Neurosci 17:1092–1099. CrossRef Medline

Martinez D, Gil R, Slifstein M, Hwang DR, Huang Y, Perez A, Kegeles L, Talbot P, Evans S, Krystal J, Laruelle M, Abi-Dargham A (2005) Alcohol dependence is associated with blunted dopamine transmission in the ventral striatum. Biol Psychiatry 58:779–786. CrossRef Medline

Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, Stephan KE (2014) Uncertainty in perception and the hierarchical Gaussian filter. Front Hum Neurosci 8:825. CrossRef Medline

Montague PR, Dolan RJ, Friston KJ, Dayan P (2012) Computational psychiatry. Trends Cogn Sci 16:72–80. CrossRef Medline

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452–454. CrossRef Medline

Park SQ, Kahnt T, Beck A, Cohen MX, Dolan RJ, Wrase J, Heinz A (2010) Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. J Neurosci 30:7749–7753. CrossRef Medline

Paulus MP, Lovero KL, Wittmann M, Leland DS (2008) Reduced behavioral and neural activation in stimulant users to different error rates during decision making. Biol Psychiatry 63:1054–1060. CrossRef Medline

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545–556. CrossRef Medline

Redish AD, Jensen S, Johnson A (2008) A unified framework for addiction: vulnerabilities in the decision process. Behav Brain Sci 31:415–437. CrossRef Medline

Reitan RM (1955) The relation of the trail making test to organic brain damage. J Consult Psychol 19:393–394. CrossRef Medline

Reiter AM, Deserno L, Wilbertz T, Heinze HJ, Schlagenhauf F (2016) Risk factors for addiction and their association with model-based behavioral control. Front Behav Neurosci 10:26.

Schlagenhauf F, Huys QJ, Deserno L, Rapp MA, Beck A, Heinze HJ, Dolan R, Heinz A (2014) Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. Neuroimage 89:171–180. CrossRef Medline

Schmidt K-H, Metzler P (1992) Wortschatztest (WST). Weinheim, Germany: Beltz Test GmbH.

Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hägele C, Keller J, Jünger E, Kathmann N, Smolka MN, Rapp MA, Schlagenhauf F, Heinz A, Huys QJ (2014) Model-based and model-free decisions in alcohol dependence. Neuropsychobiology 70:122–131. CrossRef Medline

Shohamy D, Wagner AD (2008) Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron 60: 378–389. CrossRef Medline

Sjoerds Z, de Wit S, van den Brink W, Robbins TW, Beekman AT, Penninx BW, Veltman DJ (2013) Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. Transl Psychiatry 3:e337. CrossRef Medline

Sobell LC, Sobell MB (1992) Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Measuring alcohol consumption: psychosocial and biological methods (Litten RZ, Allen JP, eds), pp 41–72. New York: Humana.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017. CrossRef Medline

Sutton RS (1992) Gain adaptation beats least squares? Paper presented at the 7th Yale Workshop on Adaptive and Learning Systems, New Haven, CT, May.

Tiffany ST, Carter BL, Singleton EG (2000) Challenges in the manipulation, assessment and interpretation of craving relevant variables. Addiction 95 [Suppl 2]:S177–S187.

Trantham-Davidson H, Burnett EJ, Gass JT, Lopez MF, Mulholland PJ, Centanni SW, Floresco SB, Chandler LJ (2014) Chronic alcohol disrupts dopamine receptor activity and the cognitive function of the medial prefrontal cortex. J Neurosci 34:3706–3718. CrossRef Medline

Volkow ND, Fowler JS, Wolf AP, Schlyer D, Shiue CY, Alpert R, Dewey SL, Logan J, Bendriem B, Christman D, et al (1990) Effects of chronic cocaine abuse on postsynaptic dopamine receptors. Am J Psychiatry 147: 719–724. CrossRef Medline

Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, Enander J, Schreiber LR, Gillan C, Fineberg NA, Sahakian BJ, Robbins TW, Harrison NA, Wood J, Daw ND, Dayan P, Grant JE, Bullmore ET (2015) Disorders of compulsivity: a common bias towards learning habits. Mol Psychiatry 20:345–352. CrossRef Medline

Wechsler D (1955) Wechsler adult intelligence scale manual. New York: Psychological Corporation.

Wimmer GE, Daw ND, Shohamy D (2012) Generalization of value in reinforcement learning by humans. Eur J Neurosci 35:1092–1104. CrossRef Medline