Dorsolateral prefrontal cortex contributes to the impaired behavioral adaptation in alcohol dependence☆

Sinem Balta Beylergil[a,b,*], Anne Beck[c], Lorenz Deserno[c,d,e], Robert C. Lorenz[c,f], Michael A. Rapp[g], Florian Schlagenhauf[c,d], Andreas Heinz[c,h], Klaus Obermayer[a,b]

[a] Department of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, 10587 Berlin, Germany
[b] Bernstein Center for Computational Neuroscience Berlin, 10115 Berlin, Germany
[c] Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany
[d] Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany
[e] Department of Neurology, Otto von Guericke University, 39118 Magdeburg, Germany
[f] Center for Adaptive Rationality, Max Planck Institute for Human Development, 14195 Berlin, Germany
[g] Social and Preventive Medicine, University of Potsdam, 14469 Potsdam, Germany
[h] Cluster of Excellence NeuroCure, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

ABSTRACT

Substance-dependent individuals often lack the ability to adjust decisions flexibly in response to the changes in reward contingencies. Prediction errors (PEs) are thought to mediate flexible decision-making by updating the reward values associated with available actions. In this study, we explored whether the neurobiological correlates of PEs are altered in alcohol dependence. Behavioral, and functional magnetic resonance imaging (fMRI) data were simultaneously acquired from 34 abstinent alcohol-dependent patients (ADP) and 26 healthy controls (HC) during a probabilistic reward-guided decision-making task with dynamically changing reinforcement contingencies. A hierarchical Bayesian inference method was used to fit and compare learning models with different assumptions about the amount of task-related information subjects may have inferred during the experiment. Here, we observed that the best-fitting model was a modified Rescorla-Wagner type model, the "double-update" model, which assumes that subjects infer the knowledge that reward contingencies are anti-correlated, and integrate both actual and hypothetical outcomes into their decisions. Moreover, comparison of the best-fitting model's parameters showed that ADP were less sensitive to punishments compared to HC. Hence, decisions of ADP after punishments were loosely coupled with the expected reward values assigned to them. A correlation analysis between the model-generated PEs and the fMRI data revealed a reduced association between these PEs and the BOLD activity in the dorsolateral prefrontal cortex (DLPFC) of ADP. A hemispheric asymmetry was observed in the DLPFC when positive and negative PE signals were analyzed separately. The right DLPFC activity in ADP showed a reduced correlation with positive PEs. On the other hand, ADP, particularly the patients with high dependence severity, recruited the left DLPFC to a lesser extent than HC for processing negative PE signals. These results suggest that the DLPFC, which has been linked to adaptive control of action selection, may play an important role in cognitive inflexibility observed in alcohol dependence when reinforcement contingencies change. Particularly, the left DLPFC may contribute to this impaired behavioral adaptation, possibly by impeding the extinction of the actions that no longer lead to a reward.

1. Introduction[1]

Alcohol has been considered as the most harmful psychoactive substance when physical, psychological, and social effects are taken together (McGinnis and Foege, 1993; Nutt et al., 2010). It can cause structural and functional changes in a network of cortical and sub-cortical structures (Beck et al., 2012; Makris et al., 2008; Moriyama et al., 2002; Ratti et al., 2002). These alterations, which partly seem to persist during abstinence (Ratti et al., 2002; Zinn et al., 2004), gradually reduce cognitive control, deteriorating individual's ability to inhibit perseverative responses and adapt to the changes in environmental contingencies. Indeed, alcohol use disorder itself can be seen as an inability to adjust responses to stimuli formerly coupled with alcohol leading to habitual, perseverative consumption patterns (Stalnaker et al., 2009).

Probabilistic reversal learning task (PRLT) has been traditionally used to assess cognitive flexibility in addiction (Izquierdo and Jentsch, 2012; Swainson et al., 2000). Experiments using PRLTs have demonstrated that various substance-dependent groups including alcohol, cocaine, and stimulant-dependent patients have difficulties adapting to reversals, i.e. abrupt changes in reward contingencies (Deserno et al., 2014; Ersche et al., 2008, 2011; Park et al., 2010). Recently, there has been an increasing interest to understand the underlying computational mechanisms of these impairments in substance use disorder through reinforcement learning (RL) models (Deserno et al., 2014; Park et al., 2010; Patzelt et al., 2014; Tanabe et al., 2013). These models are based on the idea that while individuals tend to repeat the actions leading to rewards, they tend to cease the activities that give them punishments (Sutton and Barto, 1998). They rely on a teaching signal called "prediction error," which quantifies the discrepancy between the estimated reward value of an action and the actual reward obtained by selecting that action. Learning takes place as PE updates the selected action's reward value, which then guides action selection on the next trial. More recently, it has been suggested that healthy human subjects in value-based decision-making tasks not only learn the expected reward value of the action they select, but also consider what they would have obtained if they selected the alternative action (Boorman et al., 2009, 2013; Li and Daw, 2011; Lohrenz et al., 2007; Tobia et al., 2014). The RL models accommodate this counterfactual learning via an additional fictive update rule that updates the reward expectancy of the unselected option, assuming that subjects consider the anti-correlated reward structure of the PRLT such that if one choice is likely to be rewarded, the alternative is likely to be punished (Hampton et al., 2007). In this study, using a reward-guided decision-making task with anti-correlated action-outcome contingencies that abruptly change throughout the experiment, we hypothesized that subjects would infer and incorporate this latent feature of the task structure in decision-making and integrate both actual and fictive outcomes into their decisions. Based on the recent reports showing superior model-fitting performance of these "double-update" (DU) learning models (Glaescher et al., 2009; Hampton et al., 2007; Schlagenhauf et al., 2014), we hypothesized that the DU model would fit to the behavioral data better than the standard RL models that only update the value of the selected action. Previous reports with subsets of our subjects (Deserno et al., 2014; Park et al., 2010) used the basic RL model to test their hypotheses related to the blood oxygen level dependent (BOLD) activity in the

ventral striatum (VS) that varies with PEs, which has been shown to be reliably predicted by this model (Pagnoni et al., 2002). Alternatively, our approach was to compare various learning models with different assumptions about the amount of task-related information subjects may have extracted during the experiment. Our aim was to find the model that best explains the underlying computations carried out by the subjects while adjusting their responses to abruptly changing reinforcement contingencies of the task.

The combination of RL modeling and fMRI holds promise for testing various hypotheses concerning the brain mechanisms responsible for computational processes in reward-based decision-making (Glaescher and O'Doherty, 2010). Recently, by adopting this "model-based fMRI" approach (Montague et al., 2012), neural representations of learning have been compared between substance-dependent and control groups to gain insight into the cognitive rigidity of addictive behavior. Research to date has mainly focused on the striatal impairments in reward-based decision-making (Deserno et al., 2014; Park et al., 2010) because addictive drugs seem to "hijack" the reward-related processes governed by striatal structures and evoke a pattern of behavior similar to those evoked by natural rewards (Dayan, 2009; Hyman, 2005). However, drugs also cause structural and functional changes in the PFC, especially in the DLPFC, which possibly contribute to the decline of cognitive control (Charlet et al., 2014; Goldstein et al., 2004; Loeber et al., 2009; Sullivan et al., 2000). Furthermore, reduced neural recruitment in the DLPFC has been extensively reported in various drug-dependent groups performing other tasks that require cognitive flexibility (Bolla et al., 2004; Eldreth et al., 2004; Paulus et al., 2008; Salo et al., 2009; see Goldstein and Volkow, 2011 for a review). Moreover, a previous study with a subgroup of our subjects reported abnormal signal propagation between VS and DLPFC, possibly leading to impairments in modifying and controlling behavior following reinforcement (Park et al., 2010). Based on these findings, and recent evidence on the involvement of DLPFC in PRLT (Budhani et al., 2007; Cools et al., 2002; Greening et al., 2011; Mitchell et al., 2009), the focus of this study was to elaborate on DLPFC's contribution to the inability of ADP in making flexible decisions in response to the reversals of reward contingencies. We sought to capture the neural substrates of decision-making in the PFC via a model that assumes subjects infer the unobservable (latent) reward structure of the task, which is then used to choose actions that maximize reward attained. We hypothesized that the BOLD activity in the DLPFC of ADP failing to track the PE signal derived from this model would contribute to impaired behavioral adaptation in alcohol dependence.

There is growing evidence that the human brain has distinct neural mechanisms for processing rewards and punishments (Bischoff-Grethe et al., 2009; Frank et al., 2004; Liu et al., 2007; Wrase et al., 2007; Yacubian, 2006). Furthermore, recent research suggests that these mechanisms may act differently in the case of substance use disorder (Parvaz et al., 2015; Paulus et al., 2008; Rossiter et al., 2012). The mechanisms responsible for processing punishments may be of particular interest in understanding maladaptive decision-making in alcohol use disorder because aversive consequences of alcohol use seem to be often consciously acknowledged but behaviorally ignored by abusers. Previous studies using behavioral modeling showed that the actions of substance-dependent individuals usually fail to match with the punishment expectancies attached to them (Bishara et al., 2009; Fridberg et al., 2010; Stout et al., 2004; Tanabe et al., 2013). Therefore, we hypothesized that punishments received in the current experiment would have weaker effects on the decisions of ADP compared to the decisions of HC. Negative PEs play a pivotal role in the current task as they mediate the extinction of learned actions that no longer lead to a reward when reinforcement contingencies change. In alcohol dependence, abnormal representation of these signals may contribute to the difficulties in ceasing drug-related behavior hindering the maintenance of abstinence. Therefore, one of the aims of the present study was to investigate the neural correlates of abnormal encoding of negative PEs.

[1] ACC: anterior cingulate cortex, ADP: alcohol-dependent patients, ADS: Alcohol Dependence Scale, BA: Brodmann area, BOLD: blood oxygen level dependent, DIC: deviance information criterion, DLPFC: dorsolateral prefrontal cortex, DU: double-update, FMRI (or fMRI): functional magnetic resonance imaging, FWE: family wise error, HC: healthy controls, HDI: high-density interval, HMM: hidden Markov model, IPS: intraparietal sulcus, LDH: lifetime drinking history, MCMC: Markov Chain Monte Carlo, MNI: Montreal Neurological Institute, OCDS: Obsessive Compulsive Drinking Scale, PE: prediction error, [+]PE: positive prediction error, [−]PE: negative prediction error, PRLT: probabilistic reversal learning task, RL: reinforcement learning, SU: single-update, SVC: small volume correction, VS: ventral striatum.

Table 1

Sample characteristics. ADP: alcohol-dependent patients, HC: healthy controls, FTND: Fagerstrom Test for Nicotine Dependence, EDI: the Edinburgh Handedness Inventory, LDH: Lifetime Drinking History, OCDS: Obsessive Compulsive Drinking Scale, ADS: Alcohol Dependence Scale.

	ADP (34)	HC (26)	Statistics	p
Age	44.73 ± 8.27, 23–60 years	41.92 ± 9.59, 28–61 years	$t_{58} = 1.21$	0.220
Sex	All male	All male		
Smoking	25 smokers	11 smokers	$\chi^2 = 4.75$	0.020
FTND	5 ± 2.73, 1–10	3.36 ± 2.37, 0–7	$t_{34} = -1.71$	0.100
EDI	Right-handed	Right-handed		
Verbal IQ	102.85 ± 8.92, 85–125	103.80 ± 8.93, 90–125	$t_{58} = 0.41$	0.680
LDH last year (kg)	89.10 ± 166.04, 2.10–999	5.69 ± 13.27, 0.12–68.88	$t_{58} = -2.55$	0.010
OCDS sum	17.48 ± 7.09, 4–33	2.53 ± 2.56, 0–11	$t_{58} = -7.51$	0.001
OCDS craving	8.23 ± 10.06, 0–40	28.32 ± 35.60, 0–100	$t_{58} = -2.78$	0.007
ADS	15.48 ± 7.73, 1–36	–		
Days of abstinence	17.55 ± 7.92, 7–46 days	–		

It has been shown that the severity of dependence symptoms (Doyle and Donovan, 2009) and craving for alcohol (Bottlender and Soyka, 2004) are significantly related to the ability of an alcohol-dependent individual to stay abstinent in high-risk relapse situations in which the individual should override the action "to consume" alcohol. Therefore, we reasoned that impaired negative PE signaling in ADP would be related to high dependence severity and high craving for alcohol.

2. Materials and methods

2.1. Subjects

34 abstinent ADP and 26 HC (all male) participated in the current study (see Table 1 for sample characteristics). Subjects had no other neurological or psychiatric disorder and no current drug abuse other than nicotine. All ADP were diagnosed according to the International Classification of Diseases and Related Health Problems 10th edition (World Health Organization, 2004) and Diagnostic and Statistical Manual of Mental Disorders 4th edition (American Psychiatric Association, 1994). The severity of dependence and the mean craving for alcohol were assessed with the Alcohol Dependence Scale (ADS, Skinner and Horn, 1984) and the average craving subscale of Obsessive Compulsive Drinking Scale (OCDS) (Anton, 2000). The amount of alcohol intake in the past year was evaluated with the Lifetime Drinking History (LDH) questionnaire (Skinner and Sheu, 1982). The smoking severity of the subjects was also assessed with the Fagerström Test for Nicotine Dependence (Heatherton et al., 1991). During fMRI sessions, ADP had been abstinent and were free of benzodiazepine or chlormethiazole medication for at least 1 week (> 4 half-lives). Groups did not differ on age, handedness (Oldfield, 1971) or verbal intelligence as assessed with a German vocabulary test (Schmidt and Metzler, 1992). However, there were significantly more chronic cigarette smokers in ADP than HC. All statistical analyses (including the fMRI analyses) were therefore controlled for smoking status.

The study was approved by the Ethics Committee of Charité - Universitätsmedizin Berlin and all subjects signed a written consent after all procedures were explained thoroughly.

2.2. Task description

During fMRI acquisition, subjects performed a reward-guided decision-making task with dynamically changing action-outcome contingencies (Deserno et al., 2014; Park et al., 2010; Schlagenhauf et al., 2013, 2014). On each trial, subjects had to choose one of the two abstract visual stimuli presented on a computer screen for 2 s. Following the action, the selected stimulus and its outcome—either a green smiley for reward or a red frowny for punishment—stayed on the screen for 1 s. The experiment included two runs of 100 trials separated by a short break. Trial timings were jittered by an interval of 1–6.5 s.

There were three block types with the following reward contingen-

cies: 20% left- and 80% right-, (2) 80% left- and 20% right-, and (3) 50% left- and 50% right-hand choices leading to a reward, otherwise to a punishment. Reward contingencies on the two options were fully anti-correlated, so that, for instance, when one option resulted in a reward on 80% of occasions, the other option led to a punishment 80% of the time. Subjects started the experiment with either the block type (1) or (2). Block type shifted abruptly and unpredictably to any of the randomly chosen block types after ten trials (minimum block length) when subjects chose the most highly rewarding option on 70% (50% for the 3rd block type) of the trials of an entire block. Regardless of whether this learning criterion was fulfilled, reward contingencies automatically changed after the maximum block length of 16 trials.

Subjects were instructed that the aim of the task was to learn by trial and error which of the two stimuli is better than the other, i.e. has a higher chance of winning. They were asked to adapt their behavior to possible changes in reward contingencies and win as often as possible. However, they were not informed about the exact timing of contingency changes or the reward probabilities (see Supplementary material for task instructions). Before entering the fMRI scanner, subjects were asked to perform a short version without the changes in reward contingencies to become familiar with the probabilistic nature of the task.

2.3. Statistical analysis of the behavior

The total number of correct choices and the number of blocks for which the reversal criterion was met were compared between the two groups using two-sample t-tests. Response times after rewards and punishments were compared using a 2 × 2 ANCOVA, with a between-subject factor *group*, a within-subject factor *outcome valence*, and a covariate for smoking status. We also measured the extent to which the outcome information gathered by the subjects during the previous four trials was integrated into the decisions to stay on the same option (win-stay behavior) or shift to the other option (lose-shift behavior). We then tested for between-group differences in win-stay and lose-shift behavior, which were assessed by a logistic regression analysis as explained elsewhere (den Ouden et al., 2013) and in the Supplementary material.

All standard tests in this study were performed in R 3.0.2 (R Core Team, 2013). Greenhouse–Geiser correction was used whenever the sphericity assumption was violated.

2.4. Computational modeling of the behavioral data

2.4.1. Models

We adopted a behavioral modeling approach to understand the computational processes underlying the reward-based decisions of the subjects and to explore the differences between ADP and HC in these processes. We considered three groups of computational learning models with different assumptions about the amount of task-related information subjects may have inferred during the experiment

(Schlagenhauf et al., 2014). The first group of learning models consisted of the standard Rescorla-Wagner type models (Rescorla and Wagner, 1972) in which learning is based on an error measure called "prediction error". Learning takes place as the PE (denoted as δ_t in Eq. (1)), which quantifies the discrepancy between the received outcome R_t and the expected outcome $Q_t(a_t)$, updates the expected value $Q_t(a_t)$ of the selected action at the end of each trial when R_t information is revealed (Eqs. (1) and (2)). It reinforces an action or facilitates its extinction depending on whether the obtained R_t is better (positive PE) or worse (negative PE) than the expected outcome $Q_t(a_t)$.

$$\delta_t = \rho R_t - Q_t(a_t);$$
$$\text{where } R_t = \{-1, \ 1\} \tag{1}$$

$$Q_{t+1}(a_t) = Q_t(a_t) + \alpha\delta_t \tag{2}$$

$$Q_{t+1}(a_t') = Q_t(a_t') \tag{3}$$

The effect of the reinforcement on subject's decision is represented by a free parameter called reinforcement sensitivity, as denoted by ρ in Eq. (1). Higher values of ρ magnify the differences between the option values and increase the probability of the selection of the action with higher expected value. On the other hand, lower values lead to explorative decisions that are inconsistent and independent of the reward expectancies (Stout et al., 2004). This definition of sensitivity should be distinguished from the more traditional definition as the ability to derive pleasure or displeasure from the reinforcers in the experiment.

The extent to which PE updates the expected value is determined by another free parameter called learning rate α (Eq. (2)). To study the dependence-related behavioral patterns that are specific to processing reward and punishment, we allowed the reinforcement sensitivity parameter to take two distinct values, reward sensitivity (ρ_r) or punishment sensitivity (ρ_p), according to the valence of outcome (Ito and Doya, 2009; Schlagenhauf et al., 2013). Learning rate was also allowed to take two different values depending on whether the received outcome R_t is a reward (reward learning rate α_r) or a punishment (punishment learning rate α_p).

The first group of learning models tested in this study assumes that learning can only take place through experience. Therefore, they only update the expected value of the selected action $Q_t(a_t)$, while leaving the expected value of the unselected action $Q_{t+1}(a_t')$ unchanged (Eq. (3)) (hence called "single-update" models). The second class of models, called "double-update" models, extend the single update (SU) models by also taking into account the counterfactual outcome that could have been received from the unselected action (Boorman et al., 2009, 2013; Li and Daw, 2011; Lohrenz et al., 2007; Tobia et al., 2014). Based on the idea that subjects infer and utilize the knowledge that reward contingencies on two options are fully anti-correlated, double update (DU) models update the expected values of both the selected action a_t (Eq. (2)) and the unselected action a_t' (Eq. (4)). It is important to note that after receiving a reward, the update rule for the unchosen option does not assume a certain punishment (or vice versa); but a lower probability of receiving reward, therefore a higher probability of receiving punishment as there is no other type of feedback in the experiment.

$$Q_{t+1}(a_t') = Q_t(a_t') + \alpha(-\rho R_t - Q_t(a_t')) \tag{4}$$

We generated three versions of SU and DU models with different combinations of free parameters (SU1–3, DU1–3, see Table 2). Additionally, with an additional DU model (DU4), we tested the hypothesis that fictive learning signals would not be utilized in updating of the action values as effectively as actual learning signals (Matsumoto et al., 2007). This model uses a fictive learning rate parameter, which is calculated by weighting the learning rate with an additional parameter ξ. This parameter is a fractional step size, which takes a value between 0 and 1 (Eq. (5)).

Table 2

Computational learning models. Single-update (SU), double-update (DU) models, and Hidden Markov Models (HMMs) use various combinations of free parameters. The potential scale reduction factor (PSRF) values inform about the convergence of the Markov Chain Monte Carlo (MCMC) algorithm. The minimum deviance information criterion (DIC) value (written in bold) designates the most parsimonious model. DIC values are reported for all behavioral data including (DIC$_{ALL}$) and excluding the poorly-fitted subjects (DIC$_{Fit > Chance}$). α: learning rate, ρ: reinforcement sensitivity, ξ: fictive weight, τ: transition probability, φ: outcome probability. The parameters, which take different values according to the valence of the outcome, are marked with subscripts r for reward and p for punishment.

Model	Free parameters	PSRF	DIC$_{ALL}$	DIC$_{Fit > Chance}$
SU1	α, ρ	1.03	11,178	8260
SU2	α, ρ_r, ρ_p	1.01	10,783	7888
SU3	α_r, α_p, ρ	1.26	10,819	7910
DU1	α, ρ	1.05	10,493	7588
DU2	α, ρ_r, ρ_p	1.04	**10,015**	**7141**
DU3	α_r, α_p, ρ	1.02	10,067	7184
DU4	α, ξ, ρ	1.01	10,515	7608
DU5	α, ξ, ρ_r, ρ_p	1.01	10,025	7150
HMM1	τ, φ	1.04	10,331	7461
HMM2	τ, φ_r, φ_p	1.01	10,049	7212

$$Q_{t+1}(a_t') = Q_t(a_t') + \alpha\xi(-\rho R_t - Q_t(a_t')) \tag{5}$$

In all of the SU and DU models, action probabilities $p(a_t)$ were calculated from the expected reward values of the options using the following action selection rule,

$$p(a_t = L) = \sigma(\beta\{(Q_t(L) - Q_t(R)) - c\}) \tag{6}$$

where $\sigma(z) = 1 / (1 + \exp(-z))$ is the sigmoid function. The noise temperature parameter β in Eq. (6) controls the level of stochasticity in action selection. Adjusting the reward and punishment sensitivity parameters in SU and DU models is an alternative way to modify β, which was therefore set to 1 to avoid overparameterization. The indecision point c in Eq. (6) determines the point on the sigmoid function at which both choices are equally likely to be selected. No bias was found in the choices when reward values of options were equal (paired-sample t-test, $t_{59} = -0.944$, $p = 0.348$). Hence, c was fixed to 0.

The third group of learning models implemented in this study was Hidden Markov Models (HMMs), which assume that subjects construct a state-based representation of the task via probabilities that determine contingency changes and the outcome that would arise from selecting an action (Hampton et al., 2006; Schlagenhauf et al., 2014). HMMs assign a prior belief probability to each action $b(a_t)$, which indicates the subjective belief that an action a_t is correct, i.e. associated with the higher reward contingency. At the end of each trial upon each new outcome, prior probabilities are updated to a posterior belief probability via Bayes' rule (Jordan, 1998) (see the Supplementary material for the implementation details of HMM). An important feature of HMM is that updating of the posterior belief probabilities does not involve computations of PEs. In contrast to SU and DU models, HMM uses the outcome information as an evidence to simultaneously update the belief probabilities of all possible actions. The amount of change in the prior belief made by an outcome is called "Bayesian surprise" (Itti and Baldi, 2005), which can only be computed after belief updating. On the other hand, PEs in RL models are calculated at the time of the outcome presentation and directly used in learning (Barto et al., 2013).

HMM captures the reversal nature of the task with a free parameter called transition probability (τ), which governs the transitions among the belief states. The probability with which an outcome can be obtained in a particular belief state is represented by a free parameter called outcome probability (φ). Analogous to the distinct reward and punishment sensitivities defined in the SU and DU models, the outcome probability parameter was also allowed to take two different values according to the valence of the outcome. Reward probability (φ_r) represents the likelihood of getting a reward given that subject is in a

correct belief state; whereas punishment probability (φ_p) accounts for receiving a punishment given that subject is in an incorrect belief state. We tested two versions of the HMM. The first version assumes that the chance of getting a reward from a 'correct' belief state equals to the chance of receiving a punishment from an 'incorrect' belief state. The second version allows outcome probabilities to take different values according to the valence of the outcome (see HMM1 and HMM2 in Table 2).

2.4.2. Model fitting and model comparison

Individual and group parameters were simultaneously estimated in terms of probability distributions using a hierarchical Bayesian inference method. We adopted this method because it provides a principled approach for tackling optimization problems such as numerical stability and estimations at parameter boundaries, which often occur in the modeling of choice data (Daw, 2011; Wagenmakers et al., 2008; Wetzels et al., 2010).

A Bayesian graphical model was created for each candidate learning model in JAGS (Plummer, 2003). A Markov Chain Monte Carlo (MCMC) algorithm called Gibbs sampling was used to sample from the parameter distributions of the models. Three chains of 100,000 samples were generated. To reduce autocorrelation between the MCMC samples, only every 5th sample was retained. The first 5000 samples from each chain were discarded for burn-in, leaving 19,000 samples per chain. Group prior distributions were only weakly informed to keep the estimated parameters in a reasonable range [Uniform(0,1) for learning rate, fictive weight parameter and all parameters of the HMMs; Uniform (0,20) for reward and punishment sensitivities]. To assure convergence, MCMC chains were visually analyzed for each parameter whether they stabilized at the same region of the sample space (Gelman and Rubin, 1992a, 1992b). We also calculated the Gelman-Rubin convergence and reported the potential scale reduction factor (PSRF) of each model (Gelman and Rubin, 1992a).

For selecting the best-fitting model, we calculated the model scores in deviance information criterion (DIC) (Spiegelhalter et al., 2002, see Supplementary material for details). The model with the smallest value was selected as the best-fitting model. Furthermore, to assess the level of improvement provided by the best-fitting model over the null model in predicting subject's choices, we calculated the pseudo-R^2 values for each subject, as described elsewhere (Camerer and Ho, 1999; Daw, 2011), using the posterior means of individual parameter distributions. We used pseudo-R^2 values to single out the subjects whose behavior could not be predicted by the best-fitting model significantly better than near chance level. Based on our previous reports (Schlagenhauf et al., 2014), the threshold for near chance level was set to $p \leq 0.55$ which corresponds to pseudo-$R^2 \leq 0.1375$. This threshold value was selected to make sure that our results were not confounded by poor model fits because we believe that the behavioral data of the subjects whose model-fits are close to chance level should also be treated with caution due to the probabilistic nature of model fitting. Although not reported in this article because of space limitations, we also used two other threshold values, 0.50 and 0.52, to confirm that our results were not sensitive to the selected specific threshold value.

After exclusion of these poorly-fitted subjects, we performed the model selection analysis once more to confirm that the results were not confounded by poor model fit. Model comparison was also applied separately to the choice data of HC and ADP.

2.4.3. Comparison of the model parameters between the two groups

For each model parameter, we computed the differences between the samples of the two groups (HC > ADP) at each step of the MCMC chains. We then plotted these sample differences in a histogram. The null hypothesis (H_0) was rejected when the value zero—indicating no significant group difference—fell outside the 95% high-density interval (HDI) which spanned 95% of the histograms (Kruschke, 2010).

We also examined the relationship between the parameters of the best-fitting model and clinical questionnaire scores. We used the scores on the ADS, average craving subscale of the OCDS, and the LDH to divide ADP into two subgroups at the median values. In the first analysis, we estimated and compared the model parameters of the "severely affected" (18 subjects, ADS: $15–36 \geq 15$) and the "less severe" (16 subjects, ADS: $1–14 < 15$) ADP. We used the same parameter comparison technique described above. Similarly, we categorized ADP into "high craving" (19 ADP, $OCDS_{craving}$: $10–100 \geq 10$) and "low craving" (15 subjects, $OCDS_{craving}$: $0–5 < 10$) groups using the median score on the $OCDS_{craving}$. Finally, taking the same approach, we compared the group parameters of the "high consumers" (17 ADP, $58.56–999 \, l \geq 57 \, l$) and the "low consumers" (17 ADP, $2.10–55.44 \, l < 57 \, l$), which were specified according to the median score on the LDH questionnaire. We also verified the results by testing the correlation between the posterior means of individual parameter distributions of ADP and their clinical scores.

2.4.4. Learning curves

Learning curve visualizes the adaptation of choice behavior to the reversals of reinforcement contingencies. Average learning curves of HC, ADP, and the poorly-fitted subjects were constructed by plotting the mean correct responses as a function of trial number. Choosing the stimulus with higher reward probability was considered as a correct response. The number of trials was limited to ten because blocks consisted of a minimum number of ten trials. We performed a 3×10 ANCOVA to compare the learning curves of the groups. The mean correct responses of the subjects at each trial after contingency reversals were defined as the dependent variable. The between-subjects factor *group* had three levels for HC, ADP, and the poorly-fitted subjects; whereas the within-subjects factor *trial* had ten levels for each trial after the reversals. Smoking status was included as a nuisance variable.

A successful learning model should capture and replicate the characteristics of behavioral data. To test if the best-fitting learning model fulfilled this criterion, we examined whether surrogate learning curves matched the actual learning curves of the subjects. Surrogate learning curves were constructed by plotting the mean performance of simulated data generated by letting the best-fitting model with parameters fitted to the individual subjects perform the task 100 times. Poorly-fitted data were excluded from this analysis. Surrogate data were then compared using a 2×10 group $\times$ trial ANCOVA.

2.5. FMRI data acquisition and preprocessing

Imaging was performed using a 3 Tesla GE Signa scanner with a T2*-weighted sequence (29 slices with 4 mm thickness; repetition time, 2.3 s; echo time, 27 ms; flip, 90°; matrix size, 128×128; field of view, $256 \times 256 \, mm^2$; in-plane voxel resolution of $2 \times 2 \, mm^2$) and a T1-weighted structural scan (repetition time, 7.8 ms; echo time, 3.2 ms; flip, 20°; matrix size 256×256; 1 mm slice thickness; voxel size of $1 \, mm^3$).

Functional imaging data were analyzed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). The first three volumes of each session were discarded. Volumes were corrected for the delay of slice time acquisition and motion. They were spatially normalized into MNI (Montreal Neurological Institute) space and were spatially filtered with a Gaussian kernel (8 mm full width at half maximum). Imaging data of 6 subjects (3 ADP due to motion artifacts and 3 HC due to susceptibility artifacts) were discarded. The region of interest (ROI) analyses were performed using the Marsbar toolbox in SPM (Brett et al., 2002).

2.6. FMRI data analysis

FMRI data were analyzed in an event-related manner using a general linear model approach with two levels. At the first level, reward and punishment events were modeled by stick functions at the onset of the outcome. Trial-by-trial PE time-series were computed using
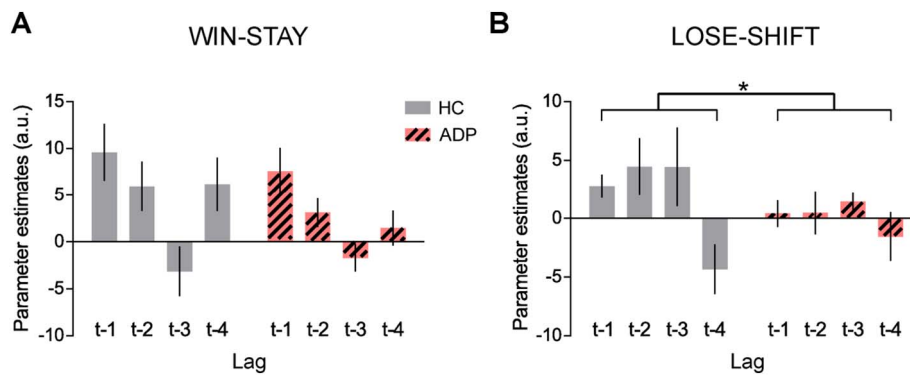
Fig. 1. Win-stay/lose-shift analysis. Parameter estimates of the (A) win-stay and (B) lose-shift regressors of the 4 trials into the past. "t" represents the time of choice. Bars denote standard errors. Asterisk denotes statistical significance (p ≤ 0.05).

the best-fitting learning model. Similar to outcome events, PE signals were also grouped into positive and negative PEs and included in the GLM as parametric modulators for reward and punishment conditions. Trials without a response were modeled separately. All regressors were convolved with the canonical hemodynamic response function as provided by SPM. The movement parameters from the realignment process were included as regressors of no interest.

2.6.1. Reward and punishment representations in the brain

"Reward vs. punishment" and "punishment vs. reward" contrast images were generated for each subject and taken to the second level. For each contrast image, we performed a random-effects group-level analysis with a one-sample *t*-test across the entire sample. We also compared groups with a two-sample *t*-test. FMRI results were reported as significant at p ≤ 0.05 family wise error (FWE) whole-brain corrected at the voxel level.

2.6.2. Neural correlates of the reward and punishment sensitivities

To investigate how reinforcement sensitivity parameters estimated in behavioral modeling were correlated with reward- and punishment-related BOLD responses across all subjects, we used two independent linear regression models at the second level of the fMRI analysis. The first regression model included the "reward vs. punishment" contrast images as the dependent variable and the reward sensitivities of the subjects as the covariate of interest. Likewise, the second regression model included the "punishment vs. reward" contrast images as the dependent variable, and the punishment sensitivities of the subjects as the covariate of interest. Results were reported significant at $p < 0.05$ SVC within the brain regions, which showed significant "reward vs. punishment" activity (for reward sensitivity), or "punishment vs. reward" activity (for punishment sensitivity) across all subjects at $p < 0.05$ FWE whole brain corrected.

2.6.3. Neural correlates of prediction errors

We performed a parametric model-based fMRI analysis to examine the differences between HC and ADP in the neural correlates of PEs. PEs were calculated using the mean values of the posterior parameter distributions estimated for each individual. Single subject contrast images of the parametric modulators positive PE ([+]PE) and negative PE ([−]PE) were taken to a 2 × 2 repeated measures ANOVA (flexible factorial design in SPM) with a between-subjects factor *group* (HC vs. ADP) and a within-subjects factor *PE type* (positive vs. negative). *Subjects* factor in SPM was also included to model subject constants. We tested the following contrasts: (1) the PE-related activity across all subjects, (2) between-group difference in the PE-related activity, (3) between-group difference in the [+]PE-related activity, (4) between-group difference in the [−]PE-related activity. FMRI results were reported as significant at p ≤ 0.05 FWE whole-brain corrected at the voxel level.
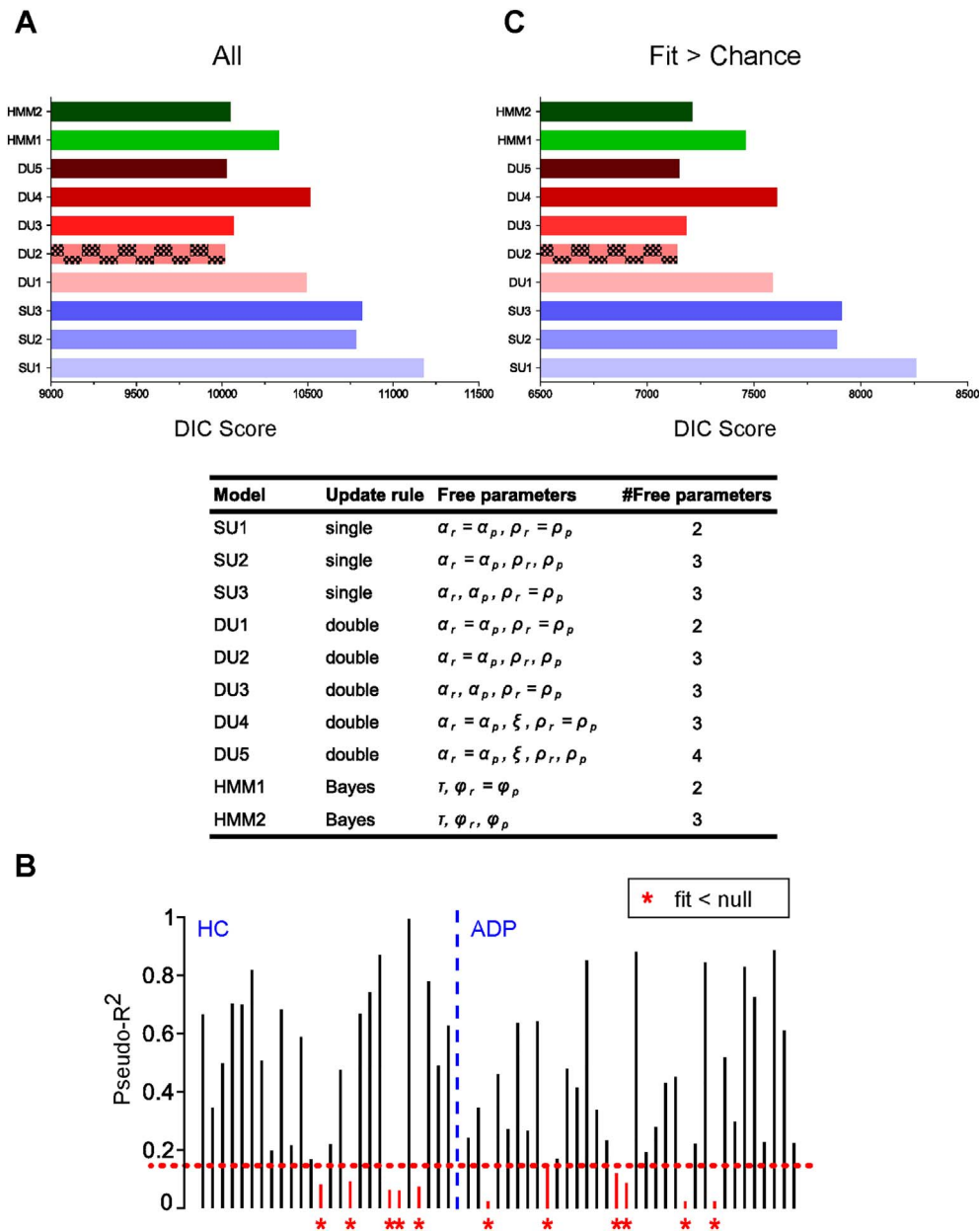
3. Results

3.1. Statistical analysis of the behavior

ADP met the learning criterion (70% correct responses during a maximum block length of 16 trials) less often than HC ($t_{58} = 1.99$, p = 0.05; $M_{HC} = 11.15$, $SD_{HC} = 3.86$; $M_{ADP} = 9.176$, $SD_{ADP} = 3.76$), completing the task with a significantly lower number of correct choices ($t_{58} = 2.586$, p = 0.012, $M_{HC} = 136.15$, $SD_{HC} = 6.14$; $M_{ADP} = 131.35$, $SD_{ADP} = 7.78$). A 2 × 2 *group × outcome valence* ANCOVA with response times as the dependent variable showed no significant main effect of *group* (F(1,58) = 0.002, p = 0.960) or *outcome valence* (F(1,58) = 2.986, p = 0.089); or a significant *group × outcome valence* interaction (F(1,58) = 1.031, p = 0.314).

We also tested for between-group differences in win-stay and lose-shift behavior. A logistic regression analysis estimated the beta parameters of the four win-stay regressors and the four lose-shift regressors. These regressors model the extent to which subjects integrated the outcome information from the previous four trials (lag) into their decisions to stay on the same option or shift to the other option. The first 2 × 4 *group × lag* ANOVA with the parameter estimates of win-stay regressors revealed no significant difference between the groups (F (1, 58) = 1.479, p = 0.229, see (Fig. 1A); however the main effect of the factor *lag* was found significant (F(3, 174) = 7.679, p < 0.0001). No significant interaction effect was found between the factors *group* and *lag* (F(3, 174) = 0.629, p = 0.597). The second 2 × 4 ANOVA with the parameter estimates of the four lose-shift regressors indicated a significant difference between HC and ADP in lose-shift behavior (F(1, 58) = 5.971, p = 0.017, see Fig. 1B). The main effect of the factor *lag* (F(3, 174) = 2.694, p < 0.047), as well as the interaction effect between the factors *group* and *lag* (F(3, 174) = 0.966, p = 0.410) were found insignificant.

3.2. Computational modeling of the behavioral data

3.2.1. Model fitting and model comparison

Potential scale reduction factors (PSRF) of the candidate models indicated that the MCMC algorithm converged for each model (see the PSRFs calculated for each model in Table 2). The model comparison analysis based on the DIC scores of the models showed that compared to the SU models, the DU models and HMMs provided superior fits to behavioral data, supporting the assumption that subjects inferred and utilized the knowledge that reward contingencies on two options are fully anti-correlated. The DU2 model (the DU model with equal reward and punishment learning rates, but distinct reward and punishment sensitivities) was selected as the best model as it fitted the behavioral data of all subjects better than the other candidate models (Fig. 2A and Table 2). The DU5 model was another candidate model with a similar DIC score. However, this model can easily be reduced to the DU2

A

All

DIC Score

C

Fit > Chance

DIC Score

Model	Update rule	Free parameters	#Free parameters
SU1	single	$\alpha_r = \alpha_p, \rho_r = \rho_p$	2
SU2	single	$\alpha_r = \alpha_p, \rho_r, \rho_p$	3
SU3	single	$\alpha_r, \alpha_p, \rho_r = \rho_p$	3
DU1	double	$\alpha_r = \alpha_p, \rho_r = \rho_p$	2
DU2	double	$\alpha_r = \alpha_p, \rho_r, \rho_p$	3
DU3	double	$\alpha_r, \alpha_p, \rho_r = \rho_p$	3
DU4	double	$\alpha_r = \alpha_p, \xi, \rho_r = \rho_p$	3
DU5	double	$\alpha_r = \alpha_p, \xi, \rho_r, \rho_p$	4
HMM1	Bayes	$\tau, \varphi_r = \varphi_p$	2
HMM2	Bayes	$\tau, \varphi_r, \varphi_p$	3

B

Fig. 2. Model comparison. (A) The DIC scores of the candidate learning models for all subjects. The most parsimonious model, the DU2 model (plotted with a patterned bar) has the lowest DIC score. (B) The pseudo-R^2 values of the subjects show the relative improvement in model-fitting provided by the DU2 model over the null model. The DU2 model was not able to predict the behavioral data of 11 subjects (5 HC and 6 ADP; marked by asterisks) better than near chance level which is marked with a horizontal dotted line at the pseudo-$R^2 = 0.1375$ (corresponding to p = 0.55). (C) The DIC scores of the candidate learning models for all subjects fitted above the near chance level. α: learning rate, ρ: reinforcement sensitivity, ξ: fictive weight, τ: transition probability, φ: outcome probability. Parameters, which take different values according to the valence of the outcome, are marked with subscripts r for reward and p for punishment.

model, because the estimated fictive weight parameter was found to be approximately equal to 1 (equal learning rates for actual and fictive outcomes).

The pseudo-R^2 values computed for each subject revealed that the DU2 model was not able to predict the behavioral data of 5 HC and 6 ADP better than near chance level (Fig. 2B). To make sure that these poorly-fitted data did not confound the model comparison results, we repeated the model selection analysis for the well-fitted subjects only. We found that the DU2 model once again explained the behavioral data significantly better than the other models (Fig. 2C and Table 2). None of the candidate models were able to predict the behavioral data of the poorly-fitted subjects better than near chance level.

Additionally, when we repeated the analysis separately for each subject group, the DU2 model provided a parsimonious fit for HC;

whereas, when only ADP were considered, the HMM2 provided a slightly better fit than the DU2 model (see Supplementary Fig. 1). HMM requires the complete model of the environment, which may be seen as a strong assumption about learning given the fact that subjects were not given the chance to practice the task beforehand (orientation version did not involve reversals). On the other hand, DU model can handle the stochastic transitions and rewards of this task without constructing the model of the environment. As a matter of fact, a comparison of the surrogate learning curves generated by these models revealed that both of these models were able to predict the behavioral data of both groups statistically alike (see Supplementary material for a comparative analysis). This similarity is also consistent with the recent studies which did not perform any model comparison analysis and used the DU model based on the assumption that this model provides a good

Table 3
Summary table of the DU2 model's estimated parameters (mean ± SD). N: sample size, α: learning rate, ρ_r: reward sensitivity, ρ_p: punishment sensitivity.

Model parameters	All subjects (N = 60)		Good-fit (N = 49)	
	HC (N = 26)	ADP (N = 34)	HC (N = 21)	ADP (N = 28)
α	0.50 ± 0.19	0.50 ± 0.30	0.51 ± 0.16	0.56 ± 0.25
ρ_r	2.19 ± 1.65	1.92 ± 1.20	2.56 ± 1.53	2.10 ± 1.27
ρ_p	1.29 ± 1.08	0.59 ± 0.61	1.52 ± 0.99	0.61 ± 0.65

approximation of the HMM for this task design while being more parsimonious (Glaescher et al., 2009; Hampton et al., 2007).

Also, our model selection was motivated by our study's hypotheses. In this study, we were particularly interested in addiction-related changes in reward-based learning guided by PEs. However, learning in HMMs do not involve computations of PE signals. Therefore, we selected the DU2 model as the best fitting model for all subjects and used this model to derive the PEs for the subsequent model-based fMRI analysis.

3.2.2. Comparison of the model parameters between the two groups

The posterior group parameter distributions of HC and ADP (see Table 3) were approximated using the DU2 model with converged MCMC samples (PSRF = 1.04, see Table 2). For each parameter of the DU2 model, parameter comparison between HC and ADP was performed by computing the differences between the samples of the two groups (HC > ADP) and plotting these differences as histograms (Fig. 3A). The null hypothesis H_0 of "no group difference (HC − ADP = 0)" was rejected only for the punishment sensitivity parameter, as the value zero fell outside the 95% HDI (0.09–1.35) of the histogram. The positive value of the mean difference

(HC − ADP = 0.705) indicates that ADP had significantly lower punishment sensitivities compared to HC. The result remained unchanged when the analysis was repeated only for the well-fitted subjects (mean difference = 0.906, 95% HDI = 0.24–1.58). On the other hand, neither the learning rates, nor the reward sensitivities showed differences between groups (learning rate: mean difference = 0.003, 95% HDI = −0.14–0.15; reward sensitivity: mean difference = 0.265, 95% HDI = −0.61–1.17).

To examine the relationship between the parameters of the DU2 model and the clinical questionnaire scores, we ran additional model fitting analyses within ADP. First, we sought to determine whether the severity of alcohol dependence (assessed with ADS) was related to the parameters of the DU2 model. The posterior parameter distributions of the less severe (LO) and the severely affected (HI) ADP were approximated using MCMC samples. Difference distributions, which were computed by subtracting the parameter distributions of the severely affected ADP from those of the less severe ADP, were plotted as difference histograms (Fig. 3B). Reward sensitivity parameter was found significantly different between these subgroups as the value zero indicating no difference was outside the 95% HDI (−2.31 − [−0.147]) of the histogram. The negative mean difference (severely affected − less severe = −1.21) indicated that the severely affected ADP had significantly higher reward sensitivities relative to the less severe ADP (M_{LO} = 1.419, SD_{LO} = 0.828; M_{HI} = 2.633, SD_{HI} = 1.672). Also, a significant positive correlation was found between the posterior means of individual reward sensitivity distributions of ADP and their ADS scores (Pearson's r = 0.482, p = 0.005).

There was no significant difference between the low craving and the high craving ADP; or between the low consumers and the high consumers.
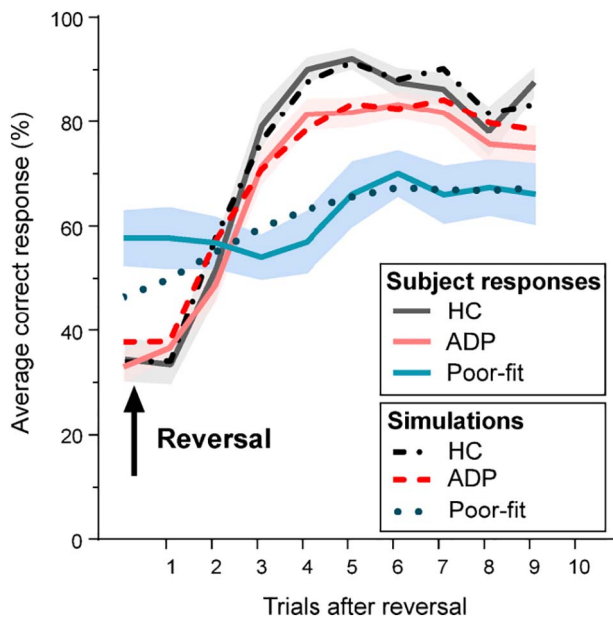
3.2.3. Learning curves

We constructed the average learning curves of HC, ADP, and the
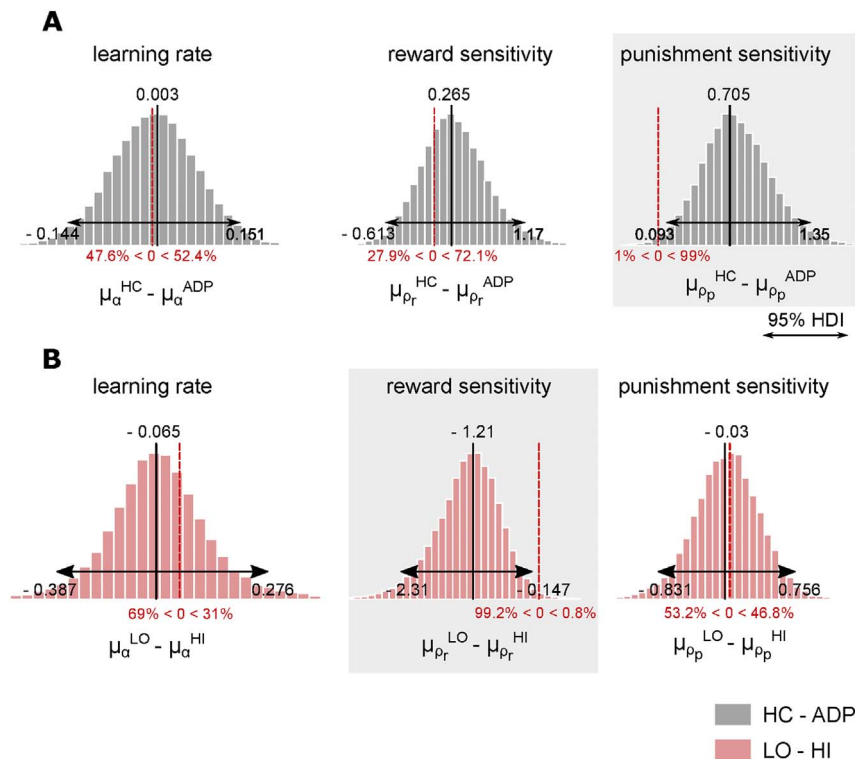
Fig. 3. Histograms of parameter differences. A. Between-group comparisons in DU2 model parameters indicate lower punishment sensitivity in ADP. B. Parameter comparison between less severe (LO) and severely affected (HI) ADP indicate greater reward sensitivity in severely affected ADP. Mean values of the histograms are shown with solid black lines. The point of no group difference is marked with a red dashed line. 95% of the distributions are found within arrows. HDI: High-density interval. μ_α: group parameter distribution for learning rate, μ_{pR}: group parameter distribution for reward sensitivity, μ_{pP}: group parameter distribution for punishment sensitivity. The figure is generated by adapting the R code originally created by Kruschke (2010).

Fig. 4. Learning curves of HC, ADP, and poorly-fitted subjects. Correct responses (selection of the stimulus with higher reward probability) were averaged over blocks of 10 trials for actual (solid lines) and simulated data (dashed lines). Individually estimated parameters of the DU2 model were used for simulations. Shaded regions denote standard errors.

poorly-fitted subjects by plotting the mean correct responses as a function of trial number (bold curves in Fig. 4). Learning curves were then compared using a 3×10 *group* $\times$ *trial* ANCOVA, which showed a significant main effect of *group* ($F(2, 57) = 6.27$, $p = 0.003$), a significant main effect of *trial* ($F(3.90, 222.64) = 46.46$, $p < 0.001$, Greenhouse–Geiser corrected) and a significant *group* $\times$ *trial* interaction ($F(7.81, 222.64) = 6.351$, $p < 0.001$, Greenhouse–Geiser corrected). When the analysis was repeated for the well-fitted HC and ADP only, the main effect of *group* ($F(1, 47) = 5.378$, $p = 0.002$) and *trial* remained significant ($F(3.21, 151.01) = 82.707$, $p < 0.001$, Greenhouse–Geiser corrected); whereas the significant *group* $\times$ *trial* interaction effect disappeared ($F(3.21, 151.01) = 1.042$, $p = 0.404$, Greenhouse–Geiser corrected). Post hoc two-sample *t*-tests revealed a significant difference between the mean correct responses of HC and ADP at the 5th trial after reversal, at which both groups reached their highest performance ($t_{47} = 2.894$, $p = 0.028$, Holm–Bonferroni corrected, $M_{HC} = 92.09\%$, $SD_{HC} = 9.27\%$; $M_{ADP} = 81.53\%$, $SD_{ADP} = 14.64\%$).

Next, we tested whether surrogate learning curves generated by the DU2 model followed the actual learning curves of the subjects. Specifically, we were interested whether the difference in the punishment sensitivities of the DU2 model (when fitted to the ADP and HC) translated into the difference in learning curves. First, we generated surrogate choice data. DU2 models with parameters fitted to the individual subjects performed the task (100 times per model). Second, we averaged the correct responses and constructed the surrogate learning curves for HC, ADP and poorly-fitted subjects (dashed curves in Fig. 4). Finally, we compared the surrogate learning curves using a 2×10 *group* $\times$ *trial* ANCOVA. Poorly-fitted data, as well as the data recorded during blocks with L: 50%–R: 50% reward contingencies, were excluded from this analysis. ANCOVA showed a significant main effect of *group* ($F(1, 47) = 6.95$, $p = 0.011$) and a significant main effect of *trial* ($F(2.26, 106.59) = 139.19$, $p < 0.001$, Greenhouse–Geiser corrected). The *group* $\times$ *trial* interaction was found to be insignificant ($F(2.26, 106.59) = 1.81$, $p = 0.064$). Post hoc *t*-tests revealed a significant group difference in the mean correct responses at the 4th trial after the reversal ($t_{47} = 3.244$, $p = 0.01$, Holm–Bonferroni corrected, $M_{HC} = 87.62\%$, $SD_{HC} = 7.99\%$; $M_{ADP} = 78.37\%$,

$SD_{ADP} = 11.06\%$) in addition to the 5th trial after the reversal ($t_{47} = 3.273$, $p = 0.01$, Holm–Bonferroni corrected, $M_{HC} = 91.43\%$, $SD_{HC} = 5.88\%$; $M_{ADP} = 83.18\%$, $SD_{ADP} = 10.35\%$). Hence, replication of the between-group difference in learning curves using the simulated data confirmed the significant association found between the decrease in the punishment sensitivity and the impaired behavioral adaptation of ADP.

Learning curve analysis was not affected by the selection of the near chance threshold as both values yielded comparable results.

3.3. FMRI analysis

3.3.1. Reward and punishment representations in the brain

Across all subjects, compared to punishments, rewards elicited a significant BOLD response in the bilateral posterior cingulate cortex, the bilateral precuneus, and the medial orbitofrontal cortex. Additionally, the left middle/superior PFC and the right putamen displayed an increased activity for reward vs. punishment (see Supplementary Fig. 2A and Supplementary Table 1). On the other hand, a significant activation in response to punishments relative to rewards was observed bilaterally in the anterior insula/inferior PFC, the dorsal anterior cingulate cortex (ACC), and the pre-SMA (see Supplementary Fig. 2B and Supplementary Table 1). Two-sample *t*-tests revealed no significant between-group difference in the reward vs. punishment or punishment vs. reward activity ($p \geq 0.001$ uncorrected).

3.3.2. Neural correlates of the reward and punishment sensitivities

We also sought to probe whether there are neural correlates of reward and punishment sensitivity parameters. A linear regression performed at the second level of the fMRI analysis, which examined the correlation between "punishment vs. reward" activity and punishment sensitivity parameter, revealed a significant positive correlation across all subjects in the right insula/inferior PFC (MNI [x y z] = [32 21 5]; k = 11; $t_{52} = 3.80$; $p_{FWE voxel (SVC)} = 0.024$; Fig. 5). On the other hand, no significant correlation was found between "reward vs. punishment" activity and reward sensitivity parameter ($p \geq 0.001$ uncorrected).

3.3.3. Neural correlates of prediction errors

Across all subjects, neural correlations of model-derived PE were found bilaterally in the VS, the middle, superior and inferior prefrontal cortices, the ACC, the midbrain, the globus pallidi, the middle temporal lobules, as well as in the left insula, the left supramarginal gyrus, the right inferior parietal lobule, the right precuneus and the right cerebellum (see Supplementary Fig. 3 and Supplementary Table 2).

Among these regions, the contrast HC > ADP showed a significant between-group difference in the PE-related activity in the bilateral DLPFC (right: MNI [x y z], [40 33 43], $t_{52} = 5.831$, $p_{FWE peak voxel (whole-brain)} = 0.005$; left: [− 41 18 53], $t_{52} = 5.488$, $p_{FWE peak voxel (whole-brain)} = 0.014$), the bilateral dorsal premotor areas (right: [25 8 63], $t_{52} = 6.081$, $p_{FWE peak voxel (whole-brain)} = 0.002$; left: [− 41 11 53], $t_{52} = 5.23$, $p_{FWE peak voxel (whole-brain)} = 0.032$), and the right intraparietal sulcus (IPS) ([42 − 62 43], $t_{52} = 6.112$, $p_{FWE peak voxel (whole-brain)} = 0.002$) (Fig. 6A and Table 4). Striatal activity related to PE did not differ between the two groups ($p \geq 0.001$ uncorrected). Furthermore, the reverse contrast, ADP > HC showed no significant difference ($p \geq 0.001$ uncorrected). In order to address the concern that group differences observed in the DLPFC might be confounded by the individual differences in the model-fits, we repeated the 2nd level analysis only for the well-fitted subjects. The differences between HC and ADP in the PE-related activity remained significant in the left and the right DLPFCs (left DLPFC: [− 23 6 43], $t_{43} = 5.15$, $p_{FWE peak voxel (whole-brain)} = 0.05$; right DLPFC: [35 38 23], $t_{43} = 5.16$, $p_{FWE peak voxel (whole-brain)} = 0.05$).

We also analyzed the effect of PE type (positive vs. negative) on the neural correlates of PE. PE was grouped into [+]PE and [−]PE
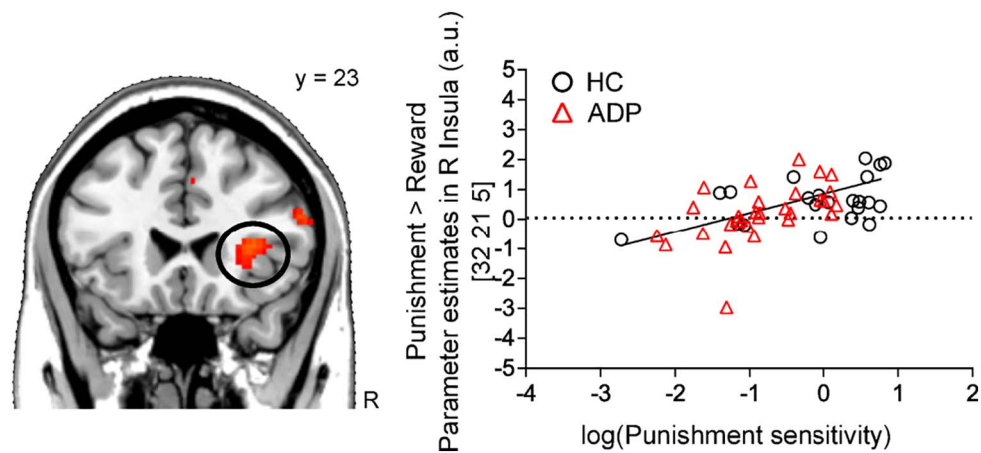
Fig. 5. Neural correlates of punishment sensitivity. The "punishment > reward" activity in the right insula is positively correlated with punishment sensitivity parameter of the best-fitting learning model. A scatter plot of the log-transformed punishment sensitivities vs. the mean parameter estimates of the punishment-related activity in the R insula (circled area) is also shown.

according to whether the obtained outcome is better ($[+]$PE) or worse ($[-]$PE) than the expected outcome. The contrast "HC > ADP" showed a hemispheric asymmetry in the DLPFC activation for the between-group differences such that a significant decrease in the $[-]$

PE-related activity ($[-38\ 11\ 53]$, $t_{52} = 5.298$, $p_{FWE\ peak\ voxel\ (whole-brain)} = 0.026$) was observed in the left DLPFC (Fig. 6B and Table 4). On the other hand, reduced $[+]$PE-related activity in ADP was found in the right DLPFC ($[40\ 33\ 40]$, $t_{52} = 5.218$, $p_{FWE\ peak\ voxel\ (whole-}$
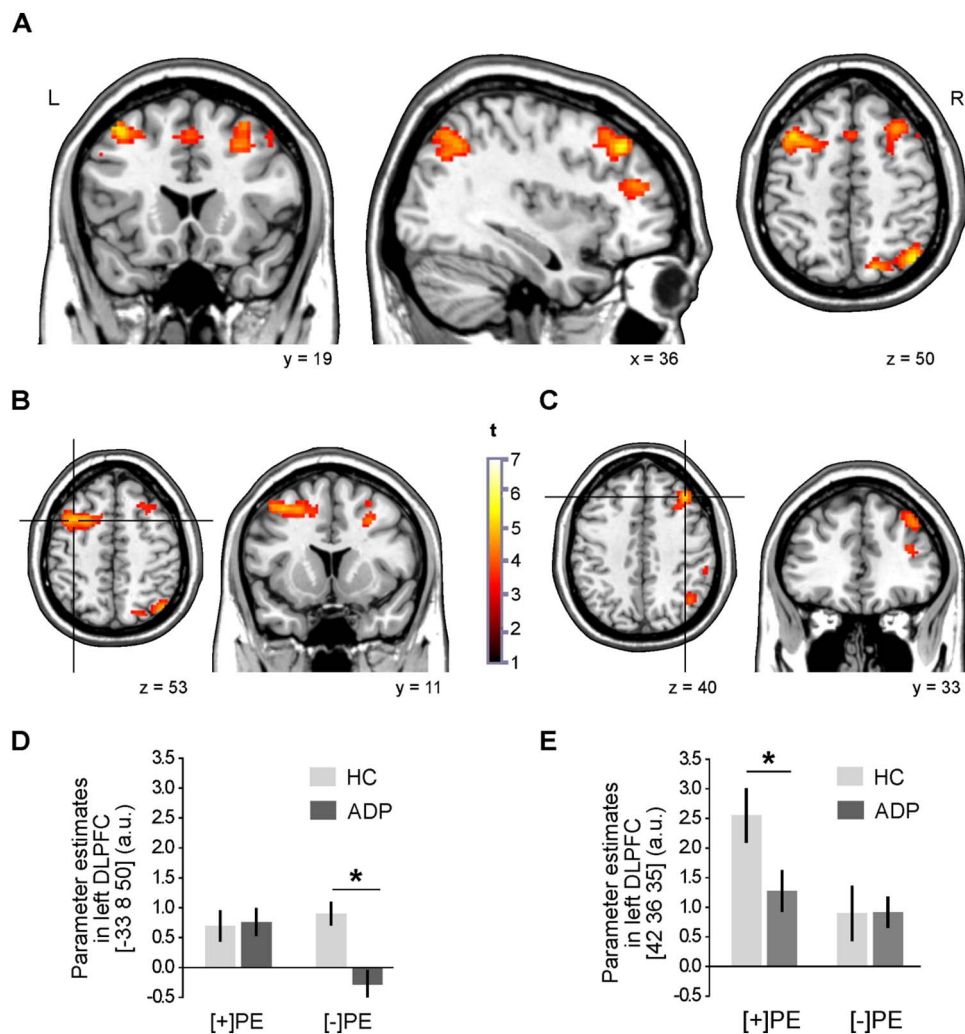
Fig. 6. Impaired PE-related activity in ADP. Group differences (HC > ADP) in the neural correlations of (A) total prediction error (PE) (both positive and negative), (B) negative prediction error ($[-]$PE), (C) positive prediction error ($[+]$PE). A threshold of p = 0.001 uncorrected with an extent threshold of 20 voxels is used for visualization (corresponds to $t > 3.31$). The color bar represents t-values. Bar plots show the beta estimates of the parametric modulators (D) $[-]$PE and (E) $[+]$PE extracted from the peak coordinates $[-33\ 8\ 50]$ and $[42\ 36\ 35]$ showing significant group $\times$ PE type interaction effect. Asterisks denote statistical significance. Error bars indicate standard errors.

Table 4
Model-based fMRI analysis results. Between-group differences (HC > ADP) in the neural correlates of the prediction error (PE), the positive PE and the negative PE. BA: Brodmann Area, k: cluster size at p < 0.001 uncorrected, FWE (whole-brain): FWE whole-brain corrected at the voxel level, MNI: Montreal Neurological Institute, HC: healthy controls, ADP: alcohol-dependent patients, PFC: prefrontal cortex, R: right, L: left.

Region	BA	k	$p_{FWE\ voxel\ (whole-brain)}$	t	MNI (x,y,z)		
PE (positive & negative) HC > ADP							
R Superior PFC	6	529	0.002	6.081	25	8	63
R Middle PFC	46		0.005	5.831	40	33	43
	9		0.028	5.280	27	23	45
L Middle PFC	9	251	0.014	5.488	−41	18	53
	9		0.032	5.230	−41	11	53
	9		0.070	4.950	−33	13	53
R Angular gyrus	39	530	0.002	6.112	42	−62	43
	7		0.073	4.934	27	−80	48
	7		0.095	4.840	17	−72	50
Positive PE HC > ADP							
R Middle PFC	46	176	0.032	5.218	40	33	40
Negative PE HC > ADP							
L Middle PFC	9	339	0.025	5.298	−38	11	53
	8		0.050	5.067	−28	11	50
R Superior PFC	6	34	0.041	5.135	25	8	65
R Angular gyrus	39	162	0.072	4.941	42	−65	45

$_{brain)}$ = 0.033) (Fig. 6C and Table 4).

This unanticipated asymmetry in the DLPFC for negative and positive PEs prompted us to perform a post hoc ANOVA. The interaction effect between *group* and *PE type* was tested using the contrasts "(HC vs. ADP) × ([−]PE vs. [+]PE)" and "(HC vs. ADP) × ([+]PE vs. [−]PE)". Results were reported as significant at p < 0.05 FWE corrected for the multiple comparisons within a volume in Brodmann area 9 and 46 that shows significant PE-related activity across all subjects. The contrast "(HC vs. ADP) × ([−]PE vs. [+]PE)" revealed a significant activation in the left DLPFC ([−33 8 50], t_{52} = 3.459, $p_{FWE\ voxel\ (SVC)}$ = 0.043, Fig. 6D and Table 5); whereas the contrast "(HC vs. ADP) × ([+]PE vs. [−]PE)" showed a significant activation in the right DLPFC ([42 36 35], t_{52} = 4.359, $p_{FWE\ voxel\ (SVC)}$ = 0.016, Fig. 6D and Table 5).

Finally, we tested whether the impairments in the [−]PE- and [+]PE-related activities in the left and right DLPFC are correlated with the clinical severity of dependence and the mean craving for alcohol as assessed with ADS and OCDS$_{craving}$, respectively. We extracted the mean parameter estimates (beta estimates) of the [−]PE- and [+]PE-related activities from the clusters showing significant group differences (cluster centers at [−38 11 53] and [40 33 40]). We found that [−]PE-related activity in the left DLPFC is significantly correlated with ADS scores of ADP (Pearson's r = −0.347, p = 0.032). This result remained

Table 5
Group × type of the prediction error (PE) interaction effects in the left and the right dorsolateral prefrontal cortices. BA: Brodmann Area, k: cluster size at p < 0.001 uncorrected, FWE voxel (SVC): FWE small volume corrected at the voxel level, MNI: Montreal Neurological Institute, HC: healthy controls, ADP: alcohol-dependent patients, [+]PE: positive prediction error, [−]PE: negative prediction error, PFC: prefrontal cortex, R: right, L: left.

Region	BA	k	$p_{FWE\ voxel\ (SVC)}$	t	MNI (x,y,z)		
Group × PE type interactions (HC vs. ADP) × ([+]PE vs. [−]PE)							
R Middle PFC	46	13	0.013	4.359	42	36	35
(HC vs. ADP) × ([−]PE vs. [+]PE)							
L Middle PFC	9	9	0.034	3.459	−33	8	50

significant when the poorly-fitted ADP were excluded from the analysis (r = −0.494, p = 0.006). No correlation was found between the [−] PE-related activity difference in the left DLPFC and OCDS$_{craving}$ scores (r = −0.001, p = 0.498). Additionally, we found that [+]PE-related activity were correlated neither with ADS (r = −0.022, p = 0.454), nor with OCDS$_{craving}$ scores (r = −0.079, p = 0.346). None of the other subscales or the total score of OCDS were correlated with the PE-related activity in the DLPFC.

4. Discussion

In this study, by using a reward-guided decision-making task and a so-called "double-update" RL model, we report a relation in alcohol dependence between impaired adaptation to the changes in reinforcement contingencies and decreased sensitivity to punishments. We also report a reduced correlation between the PEs derived from this DU model and the BOLD activity in the DLPFC of ADP. Moreover, we report an association between the severity of alcohol dependence and the decrease in the DLPFC activity related to negative PE signals, which play a critical role in adaptation to contingency changes by mediating the extinction of the behavior that is no longer associated with reward.

ADP had difficulty adapting their responses to the changing reward contingencies of the reward-guided decision-making task, a finding consistent with the results of the previous studies with subsets of our sample (13 ADP and 14 HC in Deserno et al., 2014; 20 ADP and 16 HC Park et al., 2010). Statistical analysis of win-stay and lose-shift behavior revealed that this adaptation difficulty was related to the weakened influence of punishments on decisions to shift the response. To understand the underlying computational mechanisms of this impairment, we modeled the choice behavior of our subjects using computational learning models with different assumptions about the amount of task-related information subjects may have inferred during the experiment. In line with our expectations, we found that the DU model achieved the highest accuracy in predicting the choices of all subjects. Between-group comparisons of the free parameters of this best-fitting model revealed that ADP had significantly lower punishment sensitivity. This finding is congruent with our hypothesis and the previous reports on reduced loss sensitivity and lower decision consistency in drug abuse (Ahn et al., 2014; Bishara et al., 2009; Fridberg et al., 2010; Stout et al., 2004; Tanabe et al., 2013; Vassileva et al., 2013). A computer simulation of behavioral data using the fitted parameters of the DU model reproduced the maladaptive behavior of ADP, further verifying the association between decreased punishment sensitivity and impaired behavioral adaptation. On the other hand, no significant group difference was found in other parameters of the DU model, i.e. learning rate and reward sensitivity. We argue that the difference observed between HC and ADP in adapting to changes in contingencies may not be related to learning speed or implemented learning strategy. Slower adaptation to reversals may rather be due to the fact that ADP's choices just after reversals (when subjects receive the majority of consecutive punishments) were less affected by the action values. Therefore, our results suggest that when faced with punishment, decisions of ADP are more often replaced by random guesses, which are possibly reached in the absence of deliberation. Finally, when ADP were divided into two groups at the median ADS score, we found that relative to the "less severe" group, the "severely affected" ADP had greater reward sensitivity, showing a behavioral pattern suggestive of increased tendency to respond actively to the stimuli leading to pursuit of rewards (Hyman, 2005).

Across all subjects, we discovered a positive correlation between the model-estimated punishment sensitivity and right anterior insula/inferior PFC activity in response to "punishment vs. reward". In previous neuroimaging studies featuring tasks with reversals, anterior insula, and inferior PFC responses have been shown to signal the decreases in the expected values of selected actions and predict the consecutive behavioral shifts (Cools et al., 2002; Ghahremani et al.,

2010; Glaescher et al., 2009; Hampton et al., 2006; O'Doherty et al., 2003; Schlagenhauf et al., 2014). Thus, our finding can be interpreted as evidence that the right anterior insula may be involved in the reduced ability of ADP to adjust choice behavior according to negative outcome experiences. As a part of the salient network, the anterior insula plays a crucial role in detecting salient events and engaging the central executive network for high-level cognitive control and attentional processing (see reviews by Menon and Uddin, 2010; Uddin, 2015). In our experiment, high performance partly depends on detecting the saliency of punishing stimuli, and channeling brain's top-down control resources via other cortical regions such as the DLPFC (Johnston et al., 2007). The significant correlation between punishment sensitivity and the activity in the right anterior insula therefore suggests that reduced punishment sensitivity in ADP can be related to a compromised detection of punishment events as being salient by the right anterior insula, which may fail to trigger appropriate cognitive control signals in alcohol dependence. However, it is pertinent to point out that punishment vs. reward activity in the right anterior insula/ inferior PFC did not differ between HC and ADP despite the significant difference found in punishment sensitivity. The reason for this might be that the event-related fMRI analysis per se was not able to differentiate alterations in neural activity with respect to learning or decision-making in our patient group, which motivated us to combine model-derived PEs and the fMRI data in a model-based fMRI analysis.

Model-based fMRI analysis revealed significantly lower PE-related activities in the bilateral DLPFC, the bilateral dorsal premotor areas, and the right IPS of ADP, indicating that these regions were less responsive to teaching signals that putatively facilitate behavioral adaptation. This result accords with our hypothesis that the DLPFC is implicated in the maladaptive reward-based decision-making of ADP given that the adaptive processes taking place in the PFC were captured by a computational learning model that incorporates task-related information into decisions. PEs, which form the basis for learning (Schultz and Dickinson, 2000), seem to evoke BOLD responses in the DLPFC of healthy subjects when they learned the associations between cues and affectively neutral outcomes in an associative learning task (Fletcher et al., 2001). Furthermore, Fletcher et al. demonstrated that the DLPFC activity was also able to predict the subsequent decisions of these subjects. Indeed, the tendency for taking the corrective action upon receiving an error seems to get weakened by DLPFC damage (Gehring and Knight, 2000). Similarly, transient disruption of the DLPFC activity with transcranial magnetic stimulation impairs flexible decision-making in healthy individuals (Smittenaar et al., 2013). Therefore, it is possible to interpret the observed attenuation in the PE-related DLPFC activity as a decline in ADP in the selection of the corrective action in an environment requiring adaptive responses.

To our knowledge, this is the first fMRI study with substance-dependent patients showing reduced PE-related activity in the DLPFC. Although the DLPFC has been regarded as an important neural substrate of maladaptive decision-making in substance dependence (Eldreth et al., 2004; Ersche et al., 2005; Monterosso et al., 2007; Paulus et al., 2002), a decrease in the neural tracking of PEs in this brain region has not yet been reported by other studies with substance-dependent subjects (Chiu et al., 2008; Deserno et al., 2014; Park et al., 2010; Tanabe et al., 2013). The primary reason might be that the PEs used in our model-based fMRI analysis were derived from a model that was selected from a pool of candidate models according to its performance in predicting behavioral data. On the contrary, the previous studies cited above defined the standard Rescorla–Wagner model a priori, based on their hypotheses related to the striatal PE-signaling, which has been shown to be reliably predicted by this model (Pagnoni et al., 2002). To confirm this interpretation, we repeated the model-based fMRI analysis with the PEs derived from the standard Rescorla–Wagner (denoted as "SU1" in the model set). Consistent with these studies mentioned above, we also observed significant PE-related signals in the bilateral VS (see Supplementary material). However, the

between-group difference we found in the PE-related DLPFC activity disappeared. It is probable that improvement provided by the DU model in explaining the computational processes underlying the choice behavior increased the model-based fMRI analysis's capability to capture the group differences in the neural correlates of these processes. Therefore, we conclude that selecting the learning model based on its performance on predicting behavioral data also improved the sensitivity of the subsequent model-based fMRI analysis.

Consistent with two previous studies with subsets of our subjects (Deserno et al., 2014; Park et al., 2010), we found intact striatal PE signaling in ADP, which suggests that action selection in ADP is inadequately informed by otherwise properly computed reward-learning signals in the reward/valuation network. It has been suggested that DLPFC potentiates adaptive decisions by incorporating the reward expectancies into decision representations (Barraclough et al., 2004; Christakou et al., 2009; Gold and Shadlen, 2001; Kim and Shadlen, 1999; Sugrue et al., 2005; Wallis and Miller, 2003). Consistent with this idea, simultaneous recordings from the caudate nucleus (a limbic brain structure known to encode PEs) and the lateral PFC of monkeys during a reversal learning task showed that in addition to encoding PE, the lateral PFC activity also predicts the forthcoming responses (Asaad and Eskandar, 2011). Therefore, intact striatal but reduced DLPFC activity correlated with PE suggests an ineffective integration of the reward-related information in the DLPFC of ADP which may result in selection of choices that are loosely coupled with the recently updated contingencies of the environment (Park et al., 2010; Sakagami and Watanabe, 2007).

Another way to interpret our data is that motivational signals may not be effectively embedded into cognitive processing in alcohol dependence. For reward maximization, it has recently been proposed that cognitive control function interacts with motivation (Botvinick and Braver, 2015). For instance, cognitive tasks offering monetary gains have shown that motivation can enhance executive processes to achieve efficient goal-directed behavior (e.g. Engelmann et al., 2009). Experimental data suggest that this interplay between motivation and cognition requires robust interactions between the reward/valuation network and the fronto-parietal attentional network (Pessoa, 2008; Pessoa and Engelmann, 2010). In particular, the DLPFC in the latter network appears to bridge cognitive control and value-processing by representing both cognitive and motivational (value-based) information (Dixon and Christoff, 2014). A previous report with a subset of our subject group demonstrated an abnormal functional connectivity between these two networks, specifically between the VS and the DLPFC (Park et al., 2010). Therefore, the reduced PE-related activity in the DLPFC of ADP, together with the findings of Park et al. (2010) suggest an impaired integration of motivational signals with executive control, with a possible consequence of a decrease in the engagement of cognitive control mechanisms in alcohol dependence.

The left DLPFC activity in ADP showed a decreased neural tracking of negative PEs, which, according to the RL theory, facilitate the extinction of a learned response (Schultz, 1998). This attenuated activity in the left DLPFC may contribute to the cognitive rigidity of ADP by delaying the extinction of the action that is no longer paired with a reward when reinforcement contingencies change. Diminished activity in the left DLPFC has also been demonstrated in ADP performing stop signal (Li et al., 2009) and Stroop tasks (Dao-Castellana et al., 1998), which involve extinction of "old" and reconfiguration of "new" stimulus-response associations. Moreover, transcranial magnetic stimulation of the left but not the right DLPFC disrupted the cognitive flexibility of healthy participants (Ko et al., 2008; Smittenaar et al., 2013). Here, it is important to note that the ranges of the negative PEs used in this study were determined by the punishment sensitivity parameter of the DU model estimated for each subject. Therefore, the reduced tracking of these signals suggests a neural substrate in the left DLPFC for the diminished influence of adverse consequences over the actions of ADP. An additional finding was that the right DLPFC of ADP

showed a reduced tracking of positive PEs, which may reflect impairment in initiating actions to select the action that was formerly punishing and became rewarding after a contingency reversal.

Correlating the PE-related activations in the DLPFC with a severity index of alcohol dependence (Alcohol Dependency Scale, ADS) revealed that the diminished encoding of negative PEs in the left DLPFC was more prominent in ADP with high severity scores. This finding suggests that functional abnormalities in the left DLPFC may contribute to the difficulties severely affected ADP commonly experience in overriding drug-related behavior and maintaining abstinence. On the other hand, neither the PE-related activity in the DLPFC nor the PE-related activity in the VS of ADP was found to be correlated with the craving scores of ADP (as measured using OCDS$_{craving}$). This finding is discordant with Deserno et al. (2014) showing an association between the striatal PE signals and OCDS$_{craving}$. This discrepancy, which may be due to sample-to-sample variation between these two studies or the methodological differences in behavioral modeling, needs to be clarified by future studies.

One limitation of this study was the magnetic susceptibility artifacts leading to the loss of signal intensity in the orbitofrontal cortex, as this region is located in the vicinity of the sinonasal areas. Future studies should tackle this problem with a more sensitive scanning method. Also, only male participants were recruited to avoid gender's confounding effects. Future studies with female subjects are of interest, as differences between gender groups in addictive behavior have been noted in several studies (Brady and Randall, 1999; Kosten et al., 1985; Nolen-Hoeksema, 2004). Finally, it is important to bear in mind that our correlational design limits causal inferences. Therefore, longitudinal studies are required to determine whether the alterations in the PE-related DLPFC activity reflect changes in cognitive flexibility due to alcohol dependence, or they result from preexisting vulnerabilities.

5. Conclusions

In conclusion, our results may contribute to the elucidation of the behavioral mechanisms and their neural correlates involved in impaired decision-making in substance dependence. They may, in particular, help us to understand the cognitive processes underlying the difficulties in overriding previously rewarded, but currently punishing drug-related actions with non-drug-related ones. There is some evidence that computer-aided cognitive training can treat impaired cognitive processes; improving information processing, verbal and non-verbal memory, attention, and problem-solving (Vinogradov et al., 2012). Moreover, it has successfully been shown with alcohol-dependent individuals that cognitive training can support rehabilitation as part of the traditional treatment (e.g. Fals-Stewart and Lam, 2010; Houben et al., 2011; Rupp et al., 2012). Therefore, it is possible that a focused training of adaptation to reversing reinforcement contingencies might be a valuable treatment module for improving clinical outcomes in alcohol dependence, especially in severe cases.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.nicl.2017.04.010.

References

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J.R., Kruschke, J.K., Bechara, A., Vassileva, J., 2014. Decision-making in stimulant and opiate addicts in protracted abstinence: evidence from computational modeling with pure users. Front. Psychol. 5. http://dx.doi.org/10.3389/fpsyg.2014.00849.
American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV .
Anton, R.F., 2000. Obsessive–compulsive aspects of craving: development of the obsessive compulsive drinking scale. Addiction 95, 211–217. http://dx.doi.org/10.1046/j.1360-0443.95.8s2.9.x.
Asaad, W.F., Eskandar, E.N., 2011. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. J. Neurosci. 31, 17772–17787. http://dx.doi.org/10.1523/JNEUROSCI.3793-11.2011.
Barraclough, D.J., Conroy, M.L., Lee, D., 2004. Prefrontal cortex and decision making in a mixed-strategy game. Nat. Neurosci. 7, 404–410. http://dx.doi.org/10.1038/nn1209.
Barto, A., Mirolli, M., Baldassarre, G., 2013. Novelty or surprise? Front. Psychol. 4. http://dx.doi.org/10.3389/fpsyg.2013.00907.
Beck, A., Wüstenberg, T., Genauck, A., Wrase, J., Schlagenhauf, F., Smolka, M.N., 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. http://dx.doi.org/10.1001/archgenpsychiatry.2011.2026.
Bischoff-Grethe, A., Hazeltine, E., Bergren, L., Ivry, R.B., Grafton, S.T., 2009. The influence of feedback valence in associative learning. NeuroImage 44, 243–251. http://dx.doi.org/10.1016/j.neuroimage.2008.08.038.
Bishara, A.J., Pleskac, T.J., Fridberg, D.J., Yechiam, E., Lucas, J., Busemeyer, J.R., Finn, P.R., Stout, J.C., 2009. Similar processes despite divergent behavior in two commonly used measures of risky decision making. J. Behav. Decis. Mak. 22, 435–454. http://dx.doi.org/10.1002/bdm.641.
Bolla, K.I., Ernst, M., Kiehl, K., Mouratidis, M., Eldreth, D., Contoreggi, C., Matochik, J., Kurian, V., Cadet, J., Kimes, A., Funderburk, F., London, E., 2004. Prefrontal cortical dysfunction in abstinent cocaine abusers. J. Neuropsychiatr. Clin. Neurosci. 16, 456–464. http://dx.doi.org/10.1176/jnp.16.4.456.
Boorman, E.D., Behrens, T.E.J., Woolrich, M.W., Rushworth, M.F.S., 2009. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 62, 733–743. http://dx.doi.org/10.1016/j.neuron.2009.05.014.
Boorman, E.D., Rushworth, M.F., Behrens, T.E., 2013. Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. J. Neurosci. 33, 2242–2253. http://dx.doi.org/10.1523/JNEUROSCI.3022-12.2013.
Bottlender, M., Soyka, M., 2004. Impact of craving on alcohol relapse during, and 12 months following, outpatient treatment. Alcohol Alcohol. 39, 357–361. http://dx.doi.org/10.1093/alcalc/agh073.
Botvinick, M., Braver, T., 2015. Motivation and cognitive control: from behavior to neural mechanism. Annu. Rev. Psychol. 66, 83–113. http://dx.doi.org/10.1146/annurev-psych-010814-015044.
Brady, K.T., Randall, C.L., 1999. Gender differences in substance use disorders. Psychiatr. Clin. North Am. 22, 241–252. http://dx.doi.org/10.1016/S0193-953X(05)70074-5.
Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using the MarsBar toolbox for SPM 99. NeuroImage 16, S497.
Budhani, S., Marsh, A.A., Pine, D.S., Blair, R.J.R., 2007. Neural correlates of response reversal: considering acquisition. NeuroImage 34, 1754–1765. http://dx.doi.org/10.1016/j.neuroimage.2006.08.060.
Camerer, C., Ho, T., 1999. Experience-weighted attraction learning in normal form games. Econometrica 67, 827–874. http://dx.doi.org/10.1111/1468-0262.00054.
Charlet, K., Beck, A., Jorde, A., Wimmer, L., Vollstädt-Klein, S., Gallinat, J., Walter, H., Kiefer, F., Heinz, A., 2014. Increased neural activity during high working memory load predicts low relapse risk in alcohol dependence. Addict. Biol. 19, 402–414. http://dx.doi.org/10.1111/adb.12103.
Chiu, P.H., Lohrenz, T.M., Montague, P.R., 2008. Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. Nat. Neurosci. 11, 514–520. http://dx.doi.org/10.1038/nn2067.
Christakou, A., Brammer, M., Giampietro, V., Rubia, K., 2009. Right ventromedial and dorsolateral prefrontal cortices mediate adaptive decisions under ambiguity by integrating choice utility and outcome evaluation. J. Neurosci. 29, 11020–11028. http://dx.doi.org/10.1523/JNEUROSCI.1279-09.2009.
Cools, R., Clark, L., Owen, A.M., Robbins, T.W., 2002. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J. Neurosci. 22, 4563–4567 (doi:20026435).
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R

Foundation for Statistical Computing, Vienna, Austria.

Dao-Castellana, M.H., Samson, Y., Legault, F., Martinot, J.L., Aubin, H.J., Crouzel, C., Feldman, L., Barrucand, D., Rancurel, G., Feline, A., Syrota, A., 1998. Frontal dysfunction in neurologically normal chronic alcoholic subjects: metabolic and neuropsychological findings. Psychol. Med. 28, 1039–1048.

Daw, N.D., 2011. Trial-by-trial data analysis using computational models. In: Delgado, M.R., Phelps, E.A., Robbins, T.W. (Eds.), Decision Making, Affect, and Learning: Attention and Performance XXIII, 23. Oxford University Press, Oxford, UK, pp. 3–38.

Dayan, P., 2009. Dopamine, reinforcement learning, and addiction. Pharmacopsychiatry 42, S56–S65. http://dx.doi.org/10.1055/s-0028-1124107.

Deserno, L., Beck, A., Huys, Q.J.M., Lorenz, R.C., Buchert, R., Buchholz, H.-G., Plotkin, M., Kumakara, Y., Cumming, P., Heinze, H.-J., Grace, A.A., Rapp, M.A., Schlagenhauf, F., Heinz, A., 2014. Chronic alcohol intake abolishes the relationship between dopamine synthesis capacity and learning signals in the ventral striatum. Eur. J. Neurosci. 1–10. http://dx.doi.org/10.1111/ejn.12802.

Dixon, M.L., Christoff, K., 2014. The lateral prefrontal cortex and complex value-based learning and decision making. Neurosci. Biobehav. Rev. 45, 9–18. http://dx.doi.org/10.1016/j.neubiorev.2014.04.011.

Doyle, S.R., Donovan, D.M., 2009. A validation study of the alcohol dependence scale. J. Stud. Alcohol Drugs 70, 689–699.

Eldreth, D.A., Matochik, J.A., Cadet, J.L., Bolla, K.I., 2004. Abnormal brain activity in prefrontal brain regions in abstinent marijuana users. NeuroImage 23, 914–920. http://dx.doi.org/10.1016/j.neuroimage.2004.07.032.

Engelmann, J.B., Damaraju, E., Padmala, S., Pessoa, L., 2009. Combined effects of attention and motivation on visual task performance: transient and sustained motivational effects. Front. Hum. Neurosci 3. http://dx.doi.org/10.3389/neuro.09.004.2009.

Ersche, K.D., Fletcher, P.C., Lewis, S.J.G., Clark, L., Stocks-Gee, G., London, M., Deakin, J.B., Robbins, T.W., Sahakian, B.J., 2005. Abnormal frontal activations related to decision-making in current and former amphetamine and opiate dependent individuals. Psychopharmacology 180, 612–623. http://dx.doi.org/10.1007/s00213-005-2205-7.

Ersche, K.D., Roiser, J.P., Robbins, T.W., Sahakian, B.J., 2008. Chronic cocaine but not chronic amphetamine use is associated with perseverative responding in humans. Psychopharmacology 197, 421–431. http://dx.doi.org/10.1007/s00213-007-1051-1.

Ersche, K.D., Roiser, J.P., Abbott, S., Craig, K.J., Müller, U., Suckling, J., Ooi, C., Shabbir, S.S., Clark, L., Sahakian, B.J., Fineberg, N.A., Merlo-Pich, E.V., Robbins, T.W., Bullmore, E.T., 2011. Response perseveration in stimulant dependence is associated with striatal dysfunction and can be ameliorated by a D2/3 receptor agonist. Biol. Psychiatry 70, 754–762. New Biomarkers and Treatments for Addiction. http://dx.doi.org/10.1016/j.biopsych.2011.06.033.

Fals-Stewart, W., Lam, W.K.K., 2010. Computer-assisted cognitive rehabilitation for the treatment of patients with substance use disorders: a randomized clinical trial. Exp. Clin. Psychopharmacol. 18, 87–98. http://dx.doi.org/10.1037/a0018058.

Fletcher, P.C., Anderson, J.M., Shanks, D.R., Honey, R., Carpenter, T.A., Donovan, T., Papadakis, N., Bullmore, E.T., 2001. Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. Nat. Neurosci. 4, 1043–1048. http://dx.doi.org/10.1038/nn733.

Frank, M.J., Seeberger, L.C., O'Reilly, R.C., 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940–1943. http://dx.doi.org/10.1126/science.1102941.

Fridberg, D.J., Queller, S., Ahn, W.-Y., Kim, W., Bishara, A.J., Busemeyer, J.R., Porrino, L., Stout, J.C., 2010. Cognitive mechanisms underlying risky decision-making in chronic cannabis users. J. Math. Psychol. 54, 28–38. Contributions of Mathematical Psychology to Clinical Science and Assessment. http://dx.doi.org/10.1016/j.jmp.2009.10.002.

Gehring, W.J., Knight, R.T., 2000. Prefrontal–cingulate interactions in action monitoring. Nat. Neurosci. 3, 516–520. http://dx.doi.org/10.1038/74899.

Gelman, A., Rubin, D.B., 1992a. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472.

Gelman, A., Rubin, D.B., 1992b. A single series from the Gibbs sampler provides a false sense of security. Bayesian Stat. 4, 625–631.

Ghahremani, D.G., Monterosso, J., Jentsch, J.D., Bilder, R.M., Poldrack, R.A., 2010. Neural components underlying behavioral flexibility in human reversal learning. Cereb. Cortex 20, 1843–1852. http://dx.doi.org/10.1093/cercor/bhp247.

Glaescher, J., O'Doherty, J.P., 2010. Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. Wiley Interdiscip. Rev. Cogn. Sci. 1, 501–510. http://dx.doi.org/10.1002/wcs.57.

Glaescher, J., Hampton, A.N., O'Doherty, J.P., 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb. Cortex 19, 483–495. http://dx.doi.org/10.1093/cercor/bhn098.

Gold, J.I., Shadlen, M.N., 2001. Neural computations that underlie decisions about sensory stimuli. Trends Cogn. Sci. 5, 10–16. http://dx.doi.org/10.1016/S1364-6613(00)01567-9.

Goldstein, R.Z., Volkow, N.D., 2011. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. Nat. Rev. Neurosci. 12, 652–669. http://dx.doi.org/10.1038/nrn3119.

Goldstein, R.Z., Leskovjan, A.C., Hoff, A.L., Hitzemann, R., Bashan, F., Khalsa, S.S., Wang, G.-J., Fowler, J.S., Volkow, N.D., 2004. Severity of neuropsychological impairment in cocaine and alcohol addiction: association with metabolism in the prefrontal cortex. Neuropsychologia 42, 1447–1458. http://dx.doi.org/10.1016/j.neuropsychologia.2004.04.002.

Greening, S.G., Finger, E.C., Mitchell, D.G.V., 2011. Parsing decision making processes in prefrontal cortex: response inhibition, overcoming learned avoidance, and reversal learning. NeuroImage 54, 1432–1441. http://dx.doi.org/10.1016/j.neuroimage.2010.09.017.

Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J. Neurosci. 26, 8360–8367. http://dx.doi.org/10.1523/JNEUROSCI.1010-06.2006.

Hampton, A.N., Adolphs, R., Tyszka, M.J., O'Doherty, J.P., 2007. Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. Neuron 55, 545–555. http://dx.doi.org/10.1016/j.neuron.2007.07.022.

Heatherton, T.F., Kozlowski, L.T., Frecker, R.C., Fagerstrom, K.-O., 1991. The Fagerström test for nicotine dependence: a revision of the Fagerstrom tolerance questionnaire. Br. J. Addict. 86, 1119–1127. http://dx.doi.org/10.1111/j.1360-0443.1991.tb01879.x.

Houben, K., Wiers, R.W., Jansen, A., 2011. Getting a grip on drinking behavior: training working memory to reduce alcohol abuse. Psychol. Sci. 22, 968–975. http://dx.doi.org/10.1177/0956797611412392.

Hyman, S.E., 2005. Addiction: a disease of learning and memory. Am. J. Psychiatry 162, 1414–1422. http://dx.doi.org/10.1176/appi.ajp.162.8.1414.

Ito, M., Doya, K., 2009. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. J. Neurosci. 29, 9861–9874. http://dx.doi.org/10.1523/JNEUROSCI.6157-08.2009.

Itti, L., Baldi, P.F., 2005. Bayesian surprise attracts human attention. In: Advances in Neural Information Processing Systems, pp. 547–554.

Izquierdo, A., Jentsch, J.D., 2012. Reversal learning as a measure of impulsive and compulsive behavior in addictions. Psychopharmacology 219, 607–620. http://dx.doi.org/10.1007/s00213-011-2579-7.

Johnston, K., Levin, H.M., Koval, M.J., Everling, S., 2007. Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. Neuron 53, 453–462. http://dx.doi.org/10.1016/j.neuron.2006.12.023.

Jordan, M.I., 1998. Learning in Graphical Models, Adaptive Computation and Machine Learning Series. MIT Press, Cambridge, MA, USA.

Kim, J.N., Shadlen, M.N., 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat. Neurosci. 2, 176–185. http://dx.doi.org/10.1038/5739.

Ko, J.H., Monchi, O., Ptito, A., Bloomfield, P., Houle, S., Strafella, A.P., 2008. Theta burst stimulation-induced inhibition of dorsolateral prefrontal cortex reveals hemispheric asymmetry in striatal dopamine release during a set-shifting task – a TMS-[^{11}C] raclopride PET study. Eur. J. Neurosci. 28, 2147–2155. http://dx.doi.org/10.1111/j.1460-9568.2008.06501.x.

Kosten, T.R., Rounsaville, B.J., Kleber, H.D., 1985. Ethnic and gender differences among opiate addicts. Int. J. Addict. 20, 1143–1162. http://dx.doi.org/10.3109/10826088509056356.

Kruschke, J., 2010. Doing Bayesian Data Analysis: A Tutorial Introduction with R. Academic Press, New York, NY, US.

Li, J., Daw, N.D., 2011. Signals in human striatum are appropriate for policy update rather than value prediction. J. Neurosci. 31, 5504–5511. http://dx.doi.org/10.1523/JNEUROSCI.6316-10.2011.

Li, C., Luo, X., Yan, P., Bergquist, K., Sinha, R., 2009. Altered impulse control in alcohol dependence: neural measures of stop signal performance. Alcohol. Clin. Exp. Res. 33, 740–750. http://dx.doi.org/10.1111/j.1530-0277.2008.00891.x.

Liu, X., Powell, D.K., Wang, H., Gold, B.T., Corbly, C.R., Joseph, J.E., 2007. Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. J. Neurosci. 27, 4587–4597. http://dx.doi.org/10.1523/JNEUROSCI.5227-06.2007.

Loeber, S., Duka, T., Welzel, H., Nakovics, H., Heinz, A., Flor, H., Mann, K., 2009. Impairment of cognitive abilities and decision making after chronic use of alcohol: the impact of multiple detoxifications. Alcohol Alcohol. 44, 372–381. http://dx.doi.org/10.1093/alcalc/agp030.

Lohrenz, T., McCabe, K., Camerer, C.F., Montague, P.R., 2007. Neural signature of fictive learning signals in a sequential investment task. Proc. Natl. Acad. Sci. 104, 9493–9498. http://dx.doi.org/10.1073/pnas.0608842104.

Makris, N., Oscar-Berman, M., Jaffin, S.K., Hodge, S.M., Kennedy, D.N., Caviness, V.S., Marinkovic, K., Breiter, H.C., Gasic, G.P., Harris, G.J., 2008. Decreased volume of the brain reward system in alcoholism. Biol. Psychiatry 64, 192–202. http://dx.doi.org/10.1016/j.biopsych.2008.01.018.

Matsumoto, M., Matsumoto, K., Abe, H., Tanaka, K., 2007. Medial prefrontal cell activity signaling prediction errors of action values. Nat. Neurosci. 10, 647–656. http://dx.doi.org/10.1038/nn1890.

McGinnis, J.M., Foege, W.H., 1993. Actual causes of death in the united states. JAMA 270, 2207–2212. http://dx.doi.org/10.1001/jama.1993.03510180077038.

Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 214, 655–667. http://dx.doi.org/10.1007/s00429-010-0262-0.

Mitchell, D.G.V., Luo, Q., Avny, S.B., Kasprzycki, T., Gupta, K., Chen, G., Finger, E.C., Blair, R.J.R., 2009. Adapting to dynamic stimulus-response values: differential contributions of inferior frontal, dorsomedial, and dorsolateral regions of prefrontal cortex to decision making. J. Neurosci. 29, 10827–10834. http://dx.doi.org/10.1523/JNEUROSCI.0963-09.2009.

Montague, P.R., Dolan, R.J., Friston, K.J., Dayan, P., 2012. Computational psychiatry. Trends Cogn. Sci. 16, 72–80. http://dx.doi.org/10.1016/j.tics.2011.11.018.

Monterosso, J.R., Ainslie, G., Xu, J., Cordova, X., Domier, C.P., London, E.D., 2007. Frontoparietal cortical activity of methamphetamine-dependent and comparison subjects performing a delay discounting task. Hum. Brain Mapp. 28, 383–393. http://dx.doi.org/10.1002/hbm.20281.

Moriyama, Y., Mimura, M., Kato, M., Yoshino, A., Hara, T., Kashima, H., Kato, A., Watanabe, A., 2002. Executive dysfunction and clinical outcome in chronic alcoholics. Alcohol. Clin. Exp. Res. 26, 1239–1244. http://dx.doi.org/10.1111/j.1530-0277.2002.tb02662.x.

Nolen-Hoeksema, S., 2004. Gender differences in risk factors and consequences for

alcohol use and problems. Clin. Psychol. Rev. 24, 981–1010. http://dx.doi.org/10.1016/j.cpr.2004.08.003.

Nutt, D.J., King, L.A., Phillips, L.D., 2010. Drug harms in the UK: a multicriteria decision analysis. Lancet 376, 1558–1565. http://dx.doi.org/10.1016/S0140-6736(10)61462-6.

O'Doherty, J., Critchley, H., Deichmann, R., Dolan, R.J., 2003. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. J. Neurosci. 23, 7931–7939.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113. http://dx.doi.org/10.1016/0028-3932(71)90067-4.

den Ouden, H.E.M., Daw, N.D., Fernandez, G., Elshout, J.A., Rijpkema, M., Hoogman, M., Franke, B., Cools, R., 2013. Dissociable effects of dopamine and serotonin on reversal learning. Neuron 80, 1090–1100. http://dx.doi.org/10.1016/j.neuron.2013.08.030.

Pagnoni, G., Zink, C.F., Montague, P.R., Berns, G.S., 2002. Activity in human ventral striatum locked to errors of reward prediction. Nat. Neurosci. 5, 97–98. http://dx.doi.org/10.1038/nn802.

Park, S.Q., Kahnt, T., Beck, A., Cohen, M.X., Dolan, R.J., Wrase, J., Heinz, A., 2010. Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. J. Neurosci. 30, 7749–7753. http://dx.doi.org/10.1523/JNEUROSCI.5587-09.2010.

Parvaz, M.A., Konova, A.B., Proudfit, G.H., Dunning, J.P., Malaker, P., Moeller, S.J., Maloney, T., Alia-Klein, N., Goldstein, R.Z., 2015. Impaired neural response to negative prediction errors in cocaine addiction. J. Neurosci. 35, 1872–1879. http://dx.doi.org/10.1523/JNEUROSCI.2777-14.2015.

Patzelt, E.H., Kurth-Nelson, Z., Lim, K.O., MacDonald III, A.W., 2014. Excessive state switching underlies reversal learning deficits in cocaine users. Drug Alcohol Depend. 134, 211–217. http://dx.doi.org/10.1016/j.drugalcdep.2013.09.029.

Paulus, M.P., Hozack, N.E., Zauscher, B.E., Frank, L., Brown, G.G., Braff, D.L., Schuckit, M.A., 2002. Behavioral and functional neuroimaging evidence for prefrontal dysfunction in methamphetamine-dependent subjects. Neuropsychopharmacology 26, 53–63. http://dx.doi.org/10.1016/S0893-133X(01)00334-7.

Paulus, M.P., Lovero, K.L., Wittmann, M., Leland, D.S., 2008. Reduced behavioral and neural activation in stimulant users to different error rates during decision making. Biol. Psychiatry 63, 1054–1060. http://dx.doi.org/10.1016/j.biopsych.2007.09.007.

Pessoa, L., 2008. On the relationship between emotion and cognition. Nat. Rev. Neurosci. 9, 148–158. http://dx.doi.org/10.1038/nrn2317.

Pessoa, L., Engelmann, J.B., 2010. Embedding reward signals into perception and cognition. Front. Neurosci. 4. http://dx.doi.org/10.3389/fnins.2010.00017.

Plummer, M., 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Technische Universitaet Wien, pp. 2003.

Ratti, M.T., Bo, P., Giardini, A., Soragna, D., 2002. Chronic alcoholism and the frontal lobe: which executive functions are impaired? Acta Neurol. Scand. 105, 276–281.

Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. Class. Cond. Curr. Res. Theory 64–99.

Rossiter, S., Thompson, J., Hester, R., 2012. Improving control over the impulse for reward: sensitivity of harmful alcohol drinkers to delayed reward but not immediate punishment. Drug Alcohol Depend. 125, 89–94. http://dx.doi.org/10.1016/j.drugalcdep.2012.03.017.

Rupp, C.I., Kemmler, G., Kurz, M., Hinterhuber, H., Fleischhacker, W.W., 2012. Cognitive remediation therapy during treatment for alcohol dependence. J. Stud. Alcohol Drugs 73, 625–634.

Sakagami, M., Watanabe, M., 2007. Integration of cognitive and motivational information in the primate lateral prefrontal cortex. Ann. N. Y. Acad. Sci. 1104, 89–107. http://dx.doi.org/10.1196/annals.1390.010.

Salo, R., Ursu, S., Buonocore, M.H., Leamon, M.H., Carter, C., 2009. Impaired prefrontal cortical function and disrupted adaptive cognitive control in methamphetamine abusers: a functional magnetic resonance imaging study. Biol. Psychiatry 65, 706–709. Interplay of Glutamate and Dopamine in Addiction. http://dx.doi.org/10.1016/j.biopsych.2008.11.026.

Schlagenhauf, F., Rapp, M.A., Huys, Q.J.M., Beck, A., Wüstenberg, T., Deserno, L., Buchholz, H.-G., Kalbitzer, J., Buchert, R., Bauer, M., Kienast, T., Cumming, P., Plotkin, M., Kumakura, Y., Grace, A.A., Dolan, R.J., Heinz, A., 2013. Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. Hum. Brain Mapp. 34, 1490–1499. http://dx.doi.org/10.1002/hbm.22000.

Schlagenhauf, F., Huys, Q.J.M., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.-J., Dolan, R., Heinz, A., 2014. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage 89, 171–180. http://dx.doi.org/10.1016/j.neuroimage.2013.11.034.

Schmidt, K., Metzler, P., 1992. Wortschatztest (WST). Beltz, Weinheim.

Schultz, W., 1998. Predictive reward signal of dopamine neurons. J. Neurophysiol. 80, 1–27.

Schultz, W., Dickinson, A., 2000. Neuronal coding of prediction errors. Annu. Rev. Neurosci. 23, 473–500. http://dx.doi.org/10.1146/annurev.neuro.23.1.473.

Skinner, H.A., Horn, J.L., 1984. Alcohol Dependence Scale (ADS) User's Guide. Addiction Research Foundation.

Skinner, H.A., Sheu, W.-J., 1982. Reliability of alcohol use indices; the lifetime drinking history and the MAST. J. Stud. Alcohol Drugs 43, 1157.

Smittenaar, P., FitzGerald, T.H.B., Romei, V., Wright, N.D., Dolan, R.J., 2013. Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron 80, 914–919. http://dx.doi.org/10.1016/j.neuron.2013.08.009.

Spanagel, R., Durstewitz, D., Hansson, A., Heinz, A., Kiefer, F., Köhr, G., Matthäus, F., Nöthen, M.M., Noori, H.R., Obermayer, K., Rietschel, M., Schloss, P., Scholz, H., Schumann, G., Smolka, M., Sommer, W., Vengeliene, V., Walter, H., Wurst, W., Zimmermann, U.S., Addiction GWAS Resource Group,Stringer, S., Smits, Y., Derks, E.M., 2013. A systems medicine research approach for studying alcohol addiction. Addict. Biol. 18, 883–896. http://dx.doi.org/10.1111/adb.12109.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B Stat Methodol. 64, 583–639. http://dx.doi.org/10.1111/1467-9868.00353.

Stalnaker, T.A., Takahashi, Y., Roesch, M.R., Schoenbaum, G., 2009. Neural substrates of cognitive inflexibility after chronic cocaine exposure. Neuropharmacology 56 (Supplement 1), 63–72. http://dx.doi.org/10.1016/j.neuropharm.2008.07.019.

Stout, J.C., Busemeyer, J.R., Lin, A., Grant, S.J., Bonson, K.R., 2004. Cognitive modeling analysis of decision-making processes in cocaine abusers. Psychon. Bull. Rev. 11, 742–747. http://dx.doi.org/10.3758/BF03196629.

Sugrue, L.P., Corrado, G.S., Newsome, W.T., 2005. Choosing the greater of two goods: neural currencies for valuation and decision making. Nat. Rev. Neurosci. 6, 363–375. http://dx.doi.org/10.1038/nrn1666.

Sullivan, E.V., Rosenbloom, M.J., Pfefferbaum, A., 2000. Pattern of motor and cognitive deficits in detoxified alcoholic men. Alcohol. Clin. Exp. Res. 24, 611–621. http://dx.doi.org/10.1111/j.1530-0277.2000.tb02032.x.

Sutton, R.S., Barto, A.G., 1998. Introduction to Reinforcement Learning, first ed. MIT Press, Cambridge, MA, USA.

Swainson, R., Rogers, R.D., Sahakian, B.J., Summers, B.A., Polkey, C.E., Robbins, T.W., 2000. Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. Neuropsychologia 38, 596–612.

Tanabe, J., Reynolds, J., Krmpotich, T., Claus, E., Thompson, L.L., Du, Y.P., Banich, M.T., 2013. Reduced neural tracking of prediction error in substance-dependent individuals. Am. J. Psychiatry 170, 1356–1363. http://dx.doi.org/10.1176/appi.ajp.2013.12091257.

Tobia, M.J., Guo, R., Schwarze, U., Boehmer, W., Gläscher, J., Finckh, B., Marschner, A., Büchel, C., Obermayer, K., Sommer, T., 2014. Neural systems for choice and valuation with counterfactual learning signals. NeuroImage 89, 57–69. http://dx.doi.org/10.1016/j.neuroimage.2013.11.051.

Uddin, L.Q., 2015. Salience processing and insular cortical function and dysfunction. Nat. Rev. Neurosci. 16, 55–61. http://dx.doi.org/10.1038/nrn3857.

Vassileva, J., Ahn, W.-Y., Weber, K.M., Busemeyer, J.R., Stout, J.C., Gonzalez, R., Cohen, M.H., 2013. Computational modeling reveals distinct effects of HIV and history of drug use on decision-making processes in women. PLoS One 8, e68962. http://dx.doi.org/10.1371/journal.pone.0068962.

Vinogradov, S., Fisher, M., de Villers-Sidani, E., 2012. Cognitive training for impaired neural systems in neuropsychiatric illness. Neuropsychopharmacology 37, 43–76. http://dx.doi.org/10.1038/npp.2011.251.

Wagenmakers, E.-J., Lee, M., Lodewyckx, T., Iverson, G.J., 2008. Bayesian versus Frequentist inference. In: Hoijtink, H., Klugkist, I., Boelen, P.A. (Eds.), Bayesian Evaluation of Informative Hypotheses, Statistics for Social and Behavioral Sciences. Springer New York, New York, NY, US, pp. 181–207.

Wallis, J.D., Miller, E.K., 2003. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur. J. Neurosci. 18, 2069–2081. http://dx.doi.org/10.1046/j.1460-9568.2003.02922.x.

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., Wagenmakers, E.-J., 2010. Bayesian parameter estimation in the expectancy valence model of the Iowa gambling task. J. Math. Psychol. 54, 14–27. http://dx.doi.org/10.1016/j.jmp.2008.12.001.

World Health Organization, 2004. International Statistical Classification of Diseases and Related Health Problems. World Health Organization.

Wrase, J., Kahnt, T., Schlagenhauf, F., Beck, A., Cohen, M.X., Knutson, B., Heinz, A., 2007. Different neural systems adjust motor behavior in response to reward and punishment. NeuroImage 36, 1253–1262. http://dx.doi.org/10.1016/j.neuroimage.2007.04.001.

Yacubian, J., 2006. Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. J. Neurosci. 26, 9530–9537. http://dx.doi.org/10.1523/JNEUROSCI.2915-06.2006.

Zinn, S., Stein, R., Swartzwelder, H.S., 2004. Executive functioning early in abstinence from alcohol. Alcohol. Clin. Exp. Res. 28, 1338–1346. http://dx.doi.org/10.1097/01.ALC.0000139814.81811.62.