Archival Report

When Habits Are Dangerous: Alcohol Expectancies and Habitual Decision Making Predict Relapse in Alcohol Dependence

Miriam Sebold, Stephan Nebe, Maria Garbusow, Matthias Guggenmos, Daniel J. Schad, Anne Beck, Soeren Kuitunen-Paul, Christian Sommer, Robin Frank, Peter Neu, Ulrich S. Zimmermann, Michael A. Rapp, Michael N. Smolka, Quentin J.M. Huys, Florian Schlagenhauf, and Andreas Heinz

ABSTRACT

BACKGROUND: Addiction is supposedly characterized by a shift from goal-directed to habitual decision making, thus facilitating automatic drug intake. The two-step task allows distinguishing between these mechanisms by computationally modeling goal-directed and habitual behavior as model-based and model-free control. In addicted patients, decision making may also strongly depend upon drug-associated expectations. Therefore, we investigated model-based versus model-free decision making and its neural correlates as well as alcohol expectancies in alcohol-dependent patients and healthy controls and assessed treatment outcome in patients.

METHODS: Ninety detoxified, medication-free, alcohol-dependent patients and 96 age- and gender-matched control subjects underwent functional magnetic resonance imaging during the two-step task. Alcohol expectancies were measured with the Alcohol Expectancy Questionnaire. Over a follow-up period of 48 weeks, 37 patients remained abstinent and 53 patients relapsed as indicated by the Alcohol Timeline Followback method.

RESULTS: Patients who relapsed displayed reduced medial prefrontal cortex activation during model-based decision making. Furthermore, high alcohol expectancies were associated with low model-based control in relapsers, while the opposite was observed in abstainers and healthy control subjects. However, reduced model-based control per se was not associated with subsequent relapse.

CONCLUSIONS: These findings suggest that poor treatment outcome in alcohol dependence does not simply result from a shift from model-based to model-free control but is instead dependent on the interaction between high drug expectancies and low model-based decision making. Reduced model-based medial prefrontal cortex signatures in those who relapse point to a neural correlate of relapse risk. These observations suggest that therapeutic interventions should target subjective alcohol expectancies.

Keywords: Alcohol dependence, Alcohol expectancy, Goal-directed control, Medial prefrontal cortex, Reinforcement learning, Treatment outcome

http://dx.doi.org/10.1016/j.biopsych.2017.04.019

A prominent theory in addiction research suggests that drug consumption is initially goal directed, aiming at drug-associated positive effects, then becomes habitual and eventually compulsive (1,2). This shift from goal-directed to habitual control has been suggested to be caused by long-lasting drug-associated changes in the medial prefrontal cortex (mPFC) and the ventral striatum (VS), which are involved in reward processing and reinforcement learning (3–5).

Behaviorally, there is good evidence for reduced goal-directed decision making facilitating habitual behavior in humans with substance use disorders (6), including methamphetamine (7), cocaine (8), and alcohol dependence (AD) [(9,10), but see (7)]. Overreliance on habits at the expense of goals in AD may be particularly pivotal during early abstinence,

where patients are required to inhibit automatic patterns of alcohol intake and to develop alternative coping strategies (11,12). Neuroimaging studies implicate a crucial role for the mPFC and the VS for the balance between goal-directed and habitual control (13–17), craving (18), and relapse in AD (19–21). Moreover, in animals, there is evidence that habits (e.g., automatic action tendencies) precede relapse-like behavior (22–24).

However, habit formation is not only a deficit: it is a fundamental and adaptive ability, and using habits facilitates decision making whenever cognitive resources are limited (25) or action sequences are too complex to mentally compute them (26). In AD, specific habits may be altered and induce alcohol craving, seeking, and intake. Besides habit formation,

positive alcohol expectancies as assessed by the Alcohol Expectancy Questionnaire (AEQ) (27) have been associated with current (28) and future (29,30) alcohol consumption. Explicit, self-report measures of alcohol expectancies reflect the specific expectations of the reinforcing effects of alcohol and are associated with prefrontal cortex activity and structure (31–35). One study in humans has demonstrated that acute expectation of alcohol induced by presenting alcohol beverages impairs goal-directed regulation of drug-seeking behavior in social drinkers (36), which parallels animal findings (37). Such acute expectation of alcohol may be particularly strong in subjects who have generally positive expectancies regarding the effects of alcohol consumption. Indeed, subjects who report greater positive, arousing, and social alcohol expectancies show increased appetitive responses toward alcohol cues (38). However, it is yet unclear how this association relates to real-life drinking behavior and treatment outcome in AD.

We recruited recently detoxified alcohol-dependent patients who expressed a desire to remain abstinent. We asked whether a tendency for positive alcohol expectancies interacts with model-based control and its neurobiological correlates in predicting treatment outcome.

METHODS AND MATERIALS

Participants

All data were collected as part of the Learning and Alcohol Dependence study, a bicentric German study hosted at Universitätsklinikum Dresden/Technische Universität Dresden and Charité–Universitätsmedizin Berlin. Two hundred two subjects (106 alcohol-dependent patients, 96 healthy control subjects [HCs]) completed the two-step task (39) to disentangle habitual from goal-directed decision making and the brief German version of the AEQ (27). Patients fulfilled diagnostic criteria for AD according to ICD-10 and DSM-IV-TR (40) for a minimum of 3 years. HCs were carefully matched for age, gender, education, and smoking. Exclusion criteria for all subjects were left-handedness [Edinburgh Handedness Inventory <50 (41)], a history of current or past substance use disorder (except nicotine dependence in HCs and alcohol and nicotine dependence in patients), other major psychiatric disorder [as assessed with the computer-based Composite International Diagnostic Interview (42,43)], or neurological disease. No subjects were using psychotropic medications that were known to interact with the central nervous system for at least four half-lives (including illegal drugs and detoxification treatment tested by a drug urine test). Study participation of the patients took place shortly after detoxification (Table 1). Participants gave written informed consent. Ethical approval for the study was obtained from both sites (Universitätsklinikum Dresden/Technische Universität Dresden, EK 228072012; Charité–Universitätsmedizin Berlin, EA 1/157/11), and procedures were in accordance with the Declaration of Helsinki.

Procedure

Participants were seen twice for investigation. In the first assessment, participants completed the Composite International Diagnostic Interview, a neuropsychological test battery, and additional questionnaires (Table 1). Subjects completed the German version of the AEQ at this time (27). On the second appointment, which took place shortly after the first appointment (mean ± SD, 7.0 ± 12.2 days), subjects performed the two-step task (39) along with another learning task (44). The two-step task was programmed using MATLAB software (The MathWorks, Inc., Natick, MA) with the Psychophysics Toolbox (45) and was performed while undergoing functional magnetic resonance imaging (fMRI) scanning. All participants had negative alcohol breath tests and patients were free of significant withdrawal symptoms [Clinical Institute Withdrawal Assessment of Alcohol Scale, Revised score ≤3 (46)]. Participants received compensation of 10€ an hour plus a financial bonus contingent on their performance. Blood samples for analysis of alanine transaminase, aspartate transaminase, gamma-glutamyl transferase, and phosphatidylethanol were collected.

Alcohol Expectancy Questionnaire

The brief German version of the AEQ includes 19 items. Each item describes anticipated reinforcing effects of alcohol. Items include statements such as "Alcohol generally has powerful positive effects on people (e.g., makes a person feel good or happy)" or "Alcohol helps a person to relax (e.g., feel less tense, can keep a person's mind off of mistakes at work)."

Subjects are asked to agree or disagree with each item. Disagreement and agreement of each item are coded as 1 and 2, respectively, resulting in a potential sum score between 19 and 38, for low to high expected reinforcement, respectively.

Task

Each participant performed 201 trials of the two-step task (Figure 1A for detailed task description). This task enables the analysis of model-based (goal-directed) and model-free (habitual) decisions on a trial-by-trial level, because both decision strategies make distinct predictions on choice behavior (Figure 1B).

Magnetic Resonance Imaging

fMRI was performed using a 3T Siemens Trio scanner (Siemens, Erlangen, Germany) with a 12-channel head coil. For fMRI, we used a T2-weighted echo planar imaging sequence with the following parameters: repetition time = 2410 ms, echo time = 25 ms, 80° flip angle, $3 \times 3 \times 2$ mm^3 voxel size, and a 192×192 mm^2 field of view. One volume comprised 42 transverse slices in descending order, oriented 25° to the anteroposterior commissure line. We additionally acquired a structural T1-weighted magnetization-prepared rapid gradient echo image (repetition time = 1900 ms, echo time = 2.26 ms, 9° flip angle, $1 \times 1 \times 1$ mm^3 voxel size, 256×256 mm^2 field of view).

Follow-up Procedure

After study participation, alcohol-dependent patients were regularly contacted for personal (after 4, 8, 12, 24, and 48 weeks) and telephone (after 6, 10, 18, and 36 weeks) assessments over a period of 1 year. At each contact, we assessed daily alcohol intake amount using the Alcohol Timeline

Table 1. Sample Characteristics of the Final Sample

	Group						p Values for Test Statistic			
Variable	HCs (n = 96)		Abstainers (n = 37)		Relapsers (n = 53)		Main Effect Group	HCs vs. Abstainers	Abstainers vs. Relapsers	HCs vs. Relapsers
Gender	Female: 16; male: 80		Female: 7; male: 30		Female: 6; male: 47		.56[c]	.8[c]	.37[c]	.47[c]
Site	Berlin: 56; Dresden: 40		Berlin: 24; Dresden: 13		Berlin: 28; Dresden: 25		.52[c]	.56[c]	.28[c]	.61[c]
	Mean (SD)	NA	Mean (SD)	NA	Mean (SD)	NA	F	t	t	t
Demographic Variables										
Education, years	11.9 (1.5)	2	10.8 (1.5)	2	10.6 (3.5)	2	<.05[a,d]	.2[d]	.61[d]	<.05[a,d]
Age, years	43.6 (10.9)	0	45.7 (12.0)	0	45.2 (9.9)	0	.52[b]	.36[b]	.82[b]	.38[b]
Income, €	1201 (686)	22	1150 (741)	0	1013 (621)	5	.22[d]	.61[d]	.38[d]	.08[d]
Smokers, %	65	0	75	0	75	0	.33[c]	.45[c]	1.0[c]	.45[c]
Duration of abstinence at fMRI, days	66.5 (280.9)	0	21.4 (11.6)	0	22.3 (12.4)	0	<.0001[a,d]	<.0001[a,d]	.80[d]	<.0001[a,d]
Clinical Characteristics[e]										
No. of detoxifications	—	—	2.13 (2.06)	0	4.75 (5.03)	0	<.05[a,d]	—	<.05[a]	—
Positive alcohol expectancies	25.7 (4.6)	0	31.7 (4.4)	0	32.8 (3.9)	0	<.0001[a,d]	<.0001[a,d]	.20[d]	<.0001[a,d]
Depressive symptoms	1.9 (2.3)	1	3.9 (3.9)	0	4.2 (3.7)	0	<.0001[a,d]	<.001[a,d]	.67[d]	<.0001[a,d]
Craving	2.7 (2.8)	1	10.3 (8.2)	1	12.9 (8.4)	3	<.0001[a,d]	<.0001[a,d]	.10[d]	<.0001[a,d]
Drinking motives	29 (7)	3	44 (11)	1	48 (14)	1	<.0001[a,d]	<.0001[a,d]	.36[d]	<.0001[a,d]
Time to relapse, days	—	—	—	—	87.1 (80.0)	4	—	—	—	—
Neuropsychological Testing										
Verbal IQ	28.3 (4.6)	3	28.6 (4.3)	0	28.2 (4.8)	1	.90[d]	.87[d]	.73[d]	.96[d]
Fluid IQ	10.7 (3.12)	0	9.9 (2.6)	1	9.1 (2.9)	0	<.01[a,b]	.11[b]	.26[b]	<.01[a,b]
Working memory	7.5 (2.04)	0	6.62 (1.91)	0	6.54 (1.89)	0	<.01[a,b]	<.05[a,b]	.86[b]	<.01[a,b]
Blood Markers										
AST (μKat/L)	0.45 (0.17)	28	0.69 (0.53)	5	0.71 (0.52)	11	<.001[a,d]	<.05[a,d]	.68[d]	<.001[a,d]
ALT (μKat/L)	0.43 (0.19)	28	0.88 (0.73)	5	1.08 (2.16)	11	<.001[a,d]	<.01[a,d]	.94[d]	<.001[a,d]
γ-GT (μKat/L)	0.54 (0.67)	28	3.33 (6.71)	5	1.51 (1.38)	11	<.0001[a,d]	<.0001[a,d]	.91[d]	<.0001[a,d]
PEth (ng/mL)	203.24 (359.68)	16	447.85 (349.13)	16	806.15 (736.83)	31	<.0001[a,d]	<.0001[a,d]	.14[d]	<.0001[a,d]

ALT, alanine transaminase; AST, aspartate transaminase; fMRI, functional magnetic resonance imaging; γ-GT, gamma-glutamyl transferase; HC, healthy control subjects; NA, not available; PEth, phosphatidylethanol.

[a]Significant difference.

[b]p value of linear model with group as predictor, or p value of respective contrast.

[c]p value of chi-square test.

[d]p value of Kruskal–Wallis rank sum test with group as predictor or Wilcoxon rank sum test for respective contrast.

[e]Determined as follows: positive alcohol expectancies: German version of the Alcohol Expectancy Questionnaire (71); depressive symptoms: Hospital Anxiety and Depression Scale, Subscale Depressive Symptoms (72); craving: Obsessive-Compulsive Drinking Scale (73); drinking motives were assessed using the Drinking Motives Questionnaire, revised version (52); neuropsychological testing: verbal IQ: Mehrfachwahl Wortschatz Test (74); fluid IQ: Digit Symbol Substitution Test (75); working memory: digit span backwards test from the Wechsler Adult Intelligence Scale (76).

Followback method (47), with relapse defined as consumption of 60/48 g (male/female) of alcohol on any occasion. Personal assessment included alcohol breath tests to validate self-reports. During the follow-up period, we lost 16 patients (15%). In two cases, we only had relapse reports from close relatives, which we accepted for classification. Altogether, 53 patients (59%) relapsed during the follow-up period, whereas 37 (41%) remained abstinent. Demographic and clinical characteristics of this sample are shown in Table 1.

Data Analysis

We investigated two questions: 1) whether the balance between model-free and model-based control was different between HCs and detoxified alcohol-dependent patients who remained abstinent (abstainers) and who subsequently relapsed (relapsers), and 2) whether the balance between model-free and model-based control moderated the effect of alcohol expectancies on drinking behavior. As previous studies have overwhelmingly suggested that the two-step task has power to detect variations in the goal-directed but not the habitual system (7,9,48,49), we focused on individual differences in model-based control in all analyses. We tested assumptions for all statistical analyses and computed nonparametric tests when necessary.

Task-Related Group Differences

To derive individual measurements of model-based control from behavior of the two-step task, we focused on first-stage
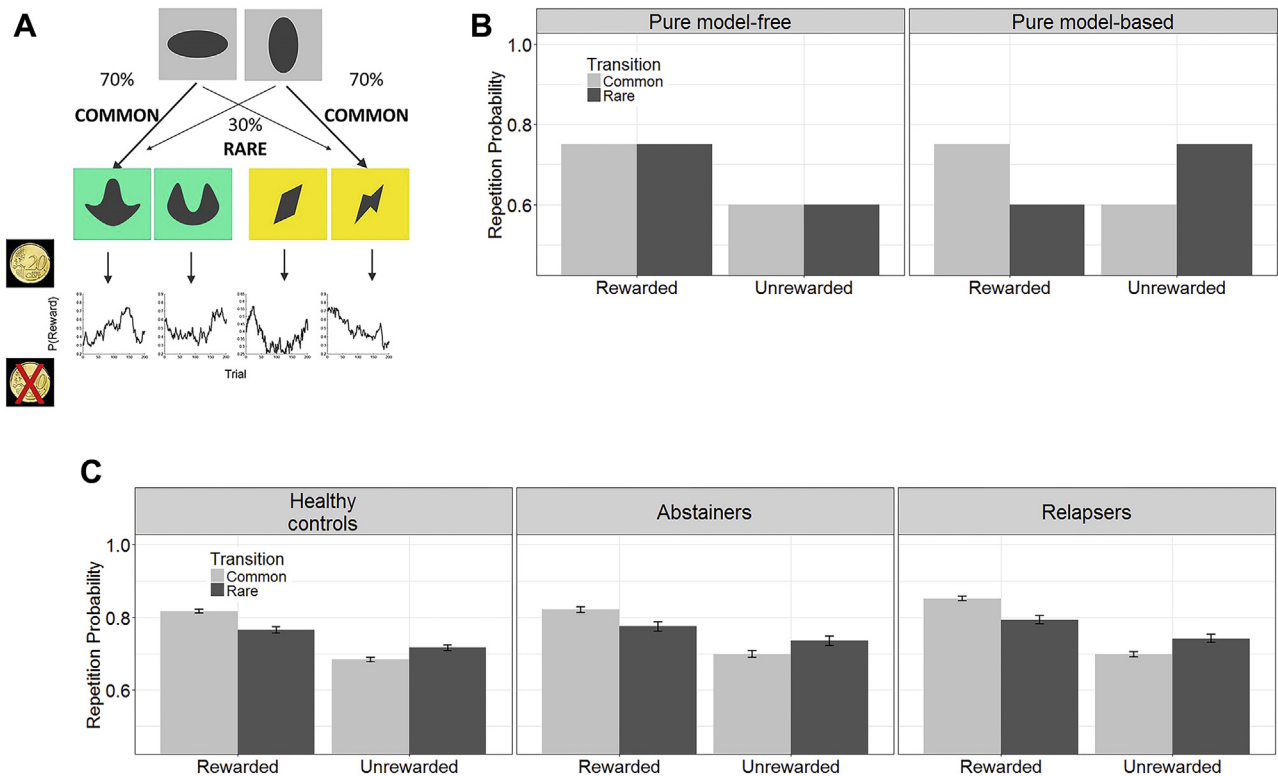
Figure 1. **(A)** An exemplary trial sequence of the two-step task. Each trial consists of two consecutive stages: participants first had to choose one of two stimuli on a gray background. This selection then led to one of two colored second-stage options (either green or yellow). Again, subjects had to choose one stimulus over the other. The transition from first-stage selections to the specific second stage was probabilistic: whereas one first stage option led frequently to the green second-stage options (70%) but rarely to the yellow second-stage options (30%), the other first-stage choice was associated with frequent yellow second-stage but rare green second-stage visits. Transition frequencies were explicitly taught during the training session with a different stimulus set. After second-stage selection, participants were probabilistically rewarded with 0.20€ or did not receive any monetary reward (0.20€ superimposed by a red X). These second-stage reward probabilities changed slowly according to Gaussian random walks with reflecting boundaries at 0.25 and 0.75 (39). In each stage, participants had 2 seconds to perform their response. Before starting the task, participants completed a training session with a different stimulus set. **(B)** Expected model-free and model-based response patterns. In pure model-free decisions, first-stage choices are repeated whenever their previous choice led to a rewarded outcome, whereas they are not repeated whenever their previous selection did not result in reward. Thus, model-free first-stage decisions are a mere function of reward from the previous trial. Contrary to this, model-based decisions take transition frequencies from first to second stage into account. For instance, in a rare trial, when a first-stage selection unexpectedly leads to a certain second-stage option and this second-stage choice then leads to reward, the best (model-based) solution to get to this rewarded second-stage choice again is to switch to the opposing first-stage choice in the next trial. **(C)** Real response pattern as a function of group. All three groups showed a mixture of model-free and model-based decision making. Groups did not differ significantly regarding their model-free or model-based choice pattern.

choices because model-free versus model-based decision making is differentially affected by reward and transition from the previous trial (39) (Figure 1B). We calculated individual model-based scores, as done previously (9), which reflect the interaction between transition frequency and reward of the previous trial (% reward common + % unrewarded rare − % rewarded rare − % unrewarded common). Model 1A involved a multinomial logistic regression analysis (multinom function from the nnet package [version 7.3-8] in R software [available at https://www.R-project.org]) to test whether group (dummy coded with three levels: HCs, abstainers, and relapsers) was predicted from model-based scores.

The raw data analysis provides a direct measurement of model-free and model-based behavior. However, it only considers trial-by-trial repetition effects. Computational models allow more comprehensive assessments, examining longer behavioral trends. Therefore, we fitted a hybrid model as previously described (39,50,51) to the behavior and estimated parameters for each subject. We used an expectation maximization algorithm to find maximum a posteriori estimates. During the fitting procedure, all subjects (HCs, abstainers, relapsers) were treated as one group.

The hybrid model contains seven parameters, of which the parameter ω is of major interest because it determines the balance between model-free (ω = 0) and model-based (ω = 1) control.

Crucially, this seven-parameter hybrid model was the best-fitting model for all groups (Supplemental Figure S1). The estimation of the parameter ω relies on the fact that subjects concurrently use model-free and model-based strategies. We excluded subjects who did not use this hybrid model as indicated by the individual log-likelihoods that did not fit better than chance (Supplement; *n* in analyses = 143). Model 1B then mirrored the analysis of the first-step repetition

probabilities: again, we performed a multinomial logistic regression analysis to test whether ω was predictive of group membership (HCs, abstainers, and relapsers).

In line with Voon *et al.* (7), we compared all other model parameters between groups (Supplemental Table S1).

Interaction Between Alcohol Expectancies and Model-Based Control

Our second hypothesis was that model-based scores would moderate the effect of alcohol expectancies on group. Model 2A tested this using multinomial logistic regression where we additionally allowed for interaction between AEQ scores and model-based control to predict group.

To elucidate the direction of our effects, we computed post hoc Spearman correlations between AEQ scores and model-based control within all groups. For illustrative purposes and further analyses, we assigned participants to high versus low alcohol expectancy groups using median splits of the AEQ (HCs median = 25; ADs median = 35).

We compared models 1A and 2A with respect to model fit. To assess the predictive capacity of the winning model, we additionally performed a cross-validation approach (stratified 10-fold cross-validation with class balancing during training).

Finally, model 2B replicated the above analysis using the computational parameter ω. We compensated for the reduced power caused by the removal of poorly fit subjects (Supplement) by using categorical AEQ information. Again, we compared models 1B and 2B with respect to model fit. Post hoc analyses were performed, comparing ω between individuals with high and low alcohol expectancies within each group using Kruskal–Wallis tests.

To evaluate whether AEQ scores were related to a motivational aspect of alcohol intake, we correlated AEQ scores with sum scores of the Drinking Motives Questionnaire (52), which measures motives of alcohol intake (53).

fMRI Analysis

Preprocessing details of the fMRI data can be found in the Supplement. All first-level analyses were based on 116 subjects (60 HCs, 21 abstainers, and 35 relapsers; see Supplemental Figure S2 for dropout details). In line with the hypothesis that relapse in AD is characterized by a shift away from model-based control, the aim of the statistical analysis of the fMRI data was to elucidate whether relapsers would show decreased model-based neural signatures in brain areas associated with the computation of these learning signals (39,50,51).

First-level analyses were conducted as previously described (39,50,51) (Supplement). Briefly, we derived individual model-free reward-prediction error (RPE_{MF}) and model-based reward-prediction error (RPE_{MB}) trajectories from the computational model under the assumption of pure model-free (ω = 0) versus full model-based control (ω = 1), respectively. In line with Daw *et al.* (39), we used means across all groups for all parameters to compute prediction errors.

Next, we used RPE_{MF} as a parametric regressor in the first-level analyses and added a second regressor—$RPE_{\Delta MB}$, the difference between RPE_{MF} and RPE_{MB}—to explain variance in

the blood oxygen level–dependent signal uniquely related to model-based prediction errors.

At the second level, contrast images for RPE_{MF} and $RPE_{\Delta MB}$ were taken to a random effects analysis. Site (Berlin vs. Dresden) was added as a covariate of no interest. For correction of multiple comparisons, familywise error (FWE) correction with p = .05 at the peak level was applied for whole brain analyses. Group comparisons in the mPFC and the VS—both areas with a pivotal role in coding RPE_{MF} and $RPE_{\Delta MB}$ signals (39,50,51,54,55)—were performed using small volume correction (SVC) with a mask containing all voxels showing a significant effect for RPE_{MF} and $RPE_{\Delta MB}$ (conjunction at p < .001 uncorrected) combining all three groups.

There is evidence for pronounced structural alterations in relapsers compared to abstainers in the mPFC, a region of interest (20,21,56). We conducted voxel-based morphometry (57) and added gray matter density as a nuisance variable in our fMRI analysis to control for morphometric alterations in the fMRI analyses (Supplemental Table S2).

To mirror the behavioral analyses, we additionally tested whether model-based neural signatures would differently correlate with AEQ scores between groups. As we had assumed that the interaction between model-based neural correlates and alcohol expectancies plays a role in the predefined regions (right/left VS and mPFC), we extracted average model-based cluster activity of these regions. Mirroring our behavioral analyses, we performed three subsequent multinomial regressions with group as dependent variable and tested for the interaction between AEQ scores and the respective cluster values.

RESULTS

Sample Characteristics

Compared to HCs, abstainers and relapsers reported significantly higher symptoms in almost all clinical characteristics, increased deficits in neuropsychological testing, and increased blood parameters related to alcohol consumption (Table 1).

Matching of HCs and alcohol-dependent patients was successful in all variables of interest (gender, school education, smoking status, and age). At baseline, there were no significant differences between abstainers and relapsers, except that the patients in the relapse group reported a larger number of previous detoxifications.

Task-Related Group Differences

Model-based control per se did not predict group membership of HCs, abstainers, or relapsers (model 1A; R^2_{McF} = .003, p = .55) (Figure 1C). The computational analysis confirmed these results. The parameter ω was not associated with group (model 1B; R^2_{McF} = .003, p = .60) (Supplemental Table S1).

Interaction Between Alcohol Expectancies and Model-Based Control

However, model-based control and alcohol expectancies interacted in predicting group membership (model 2A; R^2_{McF} = .23, p = .01). This interaction was significantly different between relapsers and HCs (p < .01) and trendwise different
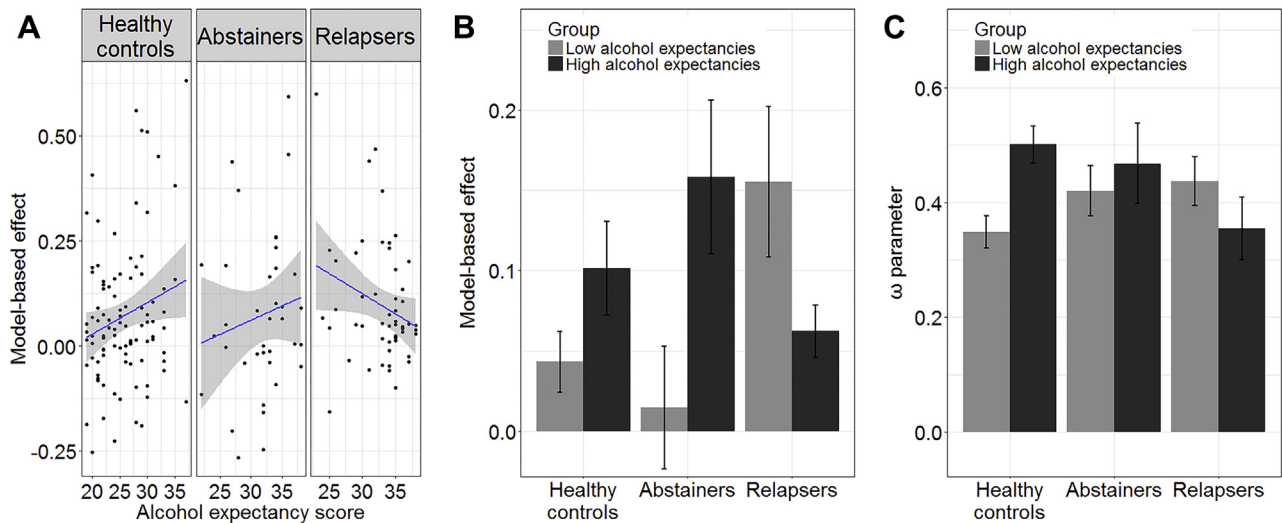
Figure 2. **(A, B)** Model-based strategy usage as a function of alcohol expectancies. Subsequent relapsers showed a negative relationship between alcohol expectancies and model-based control. This negative association was not apparent in the abstaining patients and positive in the healthy control subjects. **(C)** The relationship between ω, which indicates the balance between model-based and model-free decision making, and positive alcohol expectancies. Again, whereas healthy control subjects showed a positive association between ω and alcohol expectancies, this association was negative in relapsers and absent in abstaining patients.

between relapsers and abstainers (p = .06). Post hoc analyses using Spearman correlation to associate AEQ scores with model-based control indicated a positive association in HCs (ρ = .2, p = .04) which was absent in abstainers (p = .36, Figure 2A) and negative in relapsers (ρ = −.3, p = .03). Model comparisons between models 1A and 2A indicated that model 2A, which included the interaction between the model-based term and AEQ scores to predict group membership, outperformed model 1A, which included only the model-based term (χ^2 = 87.1, $p <$ .001). To ensure the robustness of our analysis in a predictive classification scheme, we ran the logistic regression model in a cross-validated procedure. The regression model correctly predicted group membership with an area under the curve of 0.77 (chance level: 0.5; $p < 10^{-4}$ based on a permutation test with 10,000 label permutations), corroborating the significant predictive capacity of model 2A.

Similar to our raw data analysis, model 2B indicated a significant interaction between ω and AEQ scores (R^2_{McF} = .12, p = .01), which was significantly different between relapsers and HCs (β = 1.48, $p <$ .01) and did not reach significance between relapsers and abstainers (β = 1.8, p = .1). Again, model 2B outperformed model 1B, which only included the parameter ω (χ^2 = 10.2, p = .03).

Post hoc analyses comparing high and low AEQ individuals revealed a positive association between AEQ scores and ω in HCs ($p <$ .01), but no significant association between AEQ and ω in abstainers (p = .51) and a trend toward negative association between AEQ and ω in relapsers (p = .05, Figure 2C). Adding site as a potential covariate did not change any of these results. Repeating our analyses with time to relapse as dependent variable did not reach significance (Supplement).

Among all subjects, AEQ scores were positively correlated with a variety of drinking motives (Supplemental Figures S3 and S4).

fMRI Results

Across all groups and in line with previous work (39,50,51), the conjunction between RPE$_{MF}$ and RPE$_{\Delta MB}$ reached significance in the bilateral VS (t = 6.38, x = 12, y = 12, z = −8 and t = 6.27, x = −16, y = 8, z = −10, $p_{FWE} <$ .001) and the mPFC (t = 4.85, x = −8, y = 32, z = −8, $p_{FWE} <$ .05) (Figure 3A; Supplemental Table S3). Within these regions, we found a significant correlation between neural model-based signatures (average cluster activation) and model-based scores in HCs (right VS: ρ = .29, p = .02; mPFC: ρ = .27, p = .03) (Figure 3B).

With regard to group comparisons, HCs did not differ from alcohol-dependent patients. However, with regard to treatment outcome, we observed significantly lower model-based prediction error signals (RPE$_{\Delta MB}$) in the mPFC for relapsers compared to abstainers and HCs (t = 3.9; x = −16, y = 42, z = −8, p_{FWE_SVC} = .026) (Figure 3C). Post hoc analyses, for which we extracted estimates from the peak voxel in the mPFC and compared activation between groups, indicated significantly higher activation in HCs compared to relapsers (t = 3.47, $p <$ .001) and trendwise higher activation in HCs compared to abstainers (t = 1.74, p = .08). Abstainers and relapsers did not differ (p = .10). Crucially, adding individual gray matter densities of the mPFC did not change these results (p_{FWE_SVC} = .024), suggesting that reduced neural signatures of model-based RPEs in relapsers were not caused by gray matter atrophy (Supplemental Table S2).

Model-free neural signatures did not differ between groups (Supplemental Figure S5).

Mirroring our behavioral analyses, we also examined whether AEQ scores interacted with neural correlates of model-based control in predicting group. However, the interaction between neural correlates of model-based control and AEQ scores was not significantly different between groups,
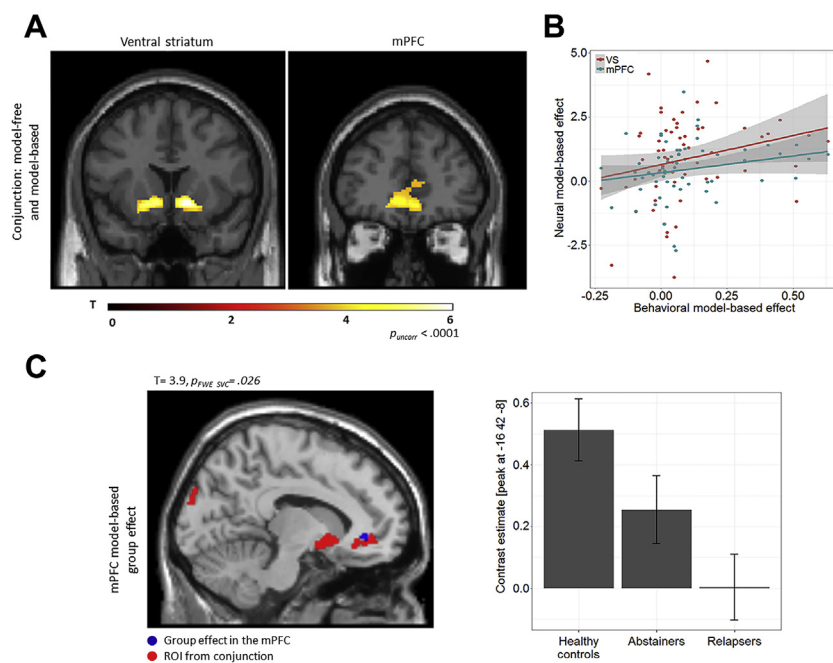
Figure 3. (A) Conjunction. Across all three groups, we found a significant coding of model-free prediction errors and additional model-based prediction errors in the ventral striatum (VS) and the medial prefrontal cortex (mPFC) (conjunction displayed at $p < .0001$ uncorrected). These regions were also the only ones that reached significance at a more conservative threshold (familywise error–corrected $p < .05$). (B) Association between neural and behavioral model-based effects. (C) Group effects. A region of the mPFC showed reduced model-based signatures for relapsers compared to abstainers and healthy control subjects. This effect survived small volume correction for the main effects of the above reported conjunction ($p_{FWE} < .026$) (panel A). Model-free signatures were not statistically different between groups. ROI, region of interest.

neither in the left (relapsers vs. abstainers, $p = .06$; relapsers vs. HCs, $p = .32$) or right VS (relapsers vs. abstainers, $p = .10$; relapsers vs. HCs, $p = .54$) nor in the mPFC (relapsers vs. abstainers, $p = .60$; relapsers vs. HCs, $p = .21$).

DISCUSSION

The main findings of our study are 1) a reduction in mPFC activation during model-based behavior in relapsers and that 2) an interaction between alcohol expectancies and goal-directed control distinguishes relapsers from abstainers and HCs. Reductions in goal-directed behavior per se were not significantly associated with AD or relapse. Instead, relapsers had high alcohol expectancies in association with low goal-directed behavior and vice versa, suggesting that the interaction between alcohol expectancies and habitual drug intake characterizes subjects with low treatment outcome.

Replicating previous studies (32,58), alcohol expectancies were correlated with drinking motives, suggesting that high alcohol expectancies reflect a motivation to consume alcohol. In abstainers and HCs, high alcohol expectancies were associated with stronger model-based control, which might help these subjects to use alcohol within a framework of self-determined values and goals. Conversely, relapsers with relatively high model-based control had low alcohol expectancies and may accordingly underestimate the effect of even low doses of alcohol to achieve a certain desired state of intoxication, whereas reductions in model-based control might facilitate excessive alcohol intake when general alcohol expectancies are high. Indeed, Hogarth et al. (36) observed that acute expectation of alcohol can temporarily interfere with goal-directed control. Our data add to this line of arguments and suggest that beyond momentary effects of alcohol expectations, a tendency to expect positive and

reinforcing alcohol effects is particularly dangerous when combined with habitual or compulsive patterns of alcohol intake (1,2). Our findings differed to some degree from a study in cocaine and polysubstance abusers, where decreased goal-directed control was found (6,8). Likewise, Voon et al. (7) observed such reduction in methamphetamine abusers but not alcohol-dependent patients, whereas a study from our own laboratory in an independent sample suggested that AD was related to reductions in goal-directed control (9). Consumption of legal drugs (e.g., alcohol) is sensitive to social traditions, including expected alcohol effects on personal well-being and social interactions. Such influences may be particularly important for subjects with AD. We also observed that functional correlates of model-based behavior in the mPFC were reduced in relapsers compared to abstainers and HCs, while at the behavioral level model-based decision making differed only between these groups when alcohol expectancies were taken into consideration. This suggests that neural activation patterns during cognitive tasks provide a valuable tool for predicting treatment outcomes (59) independent of alcohol expectancies.

Two other studies have associated blunted mPFC activation with reduced goal-directed control and flexible decision making in AD (10,60). The mPFC plays a key role in alcohol-associated behavior, including cue-induced craving in animals (61,62) and humans (63,64). Further evidence for a role of the mPFC in relapse comes from animal studies, where drug-associated mPFC activity has been shown to provoke relapse to diamorphine (65). In humans, relapse in AD has been associated with enhanced cue-related activity in the mPFC (19,20). These findings suggest that impaired mPFC function and a potential bias toward cue-induced functional activation in association with drug craving characterizes relapse across substance use disorders.

There are several limitations that need to be addressed. First, our sample size, although comparatively large, includes only a limited number of abstainers (n = 21) available for imaging, and effect sizes for the behavioral data were only moderate. Second, rodent studies have demonstrated a bias toward habitual control after chronic alcohol reward (46–48). The task here, however, used only monetary, nondrug rewards (7–10) and no alcohol cues. To what extent habitization of monetary outcomes captures the processes induced by alcohol is unclear, but ethical concerns limit the use of alcohol in detoxified subjects with AD.

Third, alcohol expectancies, although reflecting a trait rather than a state marker of motivation (66,67), are directed at consuming alcohol and are thus outcome oriented. In our study, this motivational trait was associated with low model-based control in relapsers. We do not know whether individual relapses were triggered by acute expectation of alcohol, e.g. elicited by alcohol cues. However, acute expectation of alcohol could not be tested as all subjects were motivated to remain abstinent. Additional studies in individuals with low substance use (e.g., heavy drinkers without dependence) may help to identify the effects of acute alcohol expectations on decision making.

Fourth, relapsers had gone through significantly more previous detoxifications compared to abstainers, which may contribute to neurobiological alterations associated with further and even more excessive alcohol intake, as indicated by animal experiments (68–70). However, model-based neural correlates in the mPFC were not associated with previous detoxifications in the patient group (Supplement). Finally, our study cannot disentangle preexisting conditions from alcohol-induced changes [e.g., on dopaminergic neurotransmission and its effect on goal-directed correlates (50)]; therefore, further studies employing longitudinal designs are required.

In conclusion, decreased model-based control may predict relapse only in patients with high alcohol expectancies. This study further specifies the theory of goals and habits in AD and suggests a pivotal role of alcohol expectancies, which can easily be assessed in clinical settings. Our study showed how the computational mechanism underlying goal-directed control and its neurobiological correlate (reduced mPFC activation) are associated with poor treatment outcome. The interaction between alcohol expectancies and drug taking habits points to potential therapeutic interventions that aim to increase goal-directed control (such as motivational interviewing) and alter the anticipated outcomes of alcohol use.

ACKNOWLEDGMENTS AND DISCLOSURES

REFERENCES

1. Everitt BJ, Robbins TW (2005): Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 8:1481–1489.
2. Everitt BJ, Robbins TW (2016): Drug addiction: Updating actions to habits to compulsions ten years on. Annu Rev Psychol 67:23–50.
3. Berridge KC (2012): From prediction error to incentive salience: Mesolimbic computation of reward motivation. Eur J Neurosci 35:1124–1143.
4. Volkow ND, Li TK (2004): Drug addiction: The neurobiology of behaviour gone awry. Nat Rev Neurosci 5:963–970.
5. Heinz A, Siessmeier T, Wrase J, Hermann D, Klein S, Grusser SM, et al. (2004): Correlation between dopamine D(2) receptors in the ventral striatum and central processing of alcohol cues and craving. Am J Psychiatry 161:1783–1789.
6. McKim TH, Bauer DJ, Boettiger CA (2016): Addiction history associates with the propensity to form habits. J Cogn Neurosci 28:1024–1038.
7. Voon V, Derbyshire K, Ruck C, Irvine MA, Worbe Y, Enander J, et al. (2014): Disorders of compulsivity: A common bias towards learning habits. Mol Psychiatry 20:345–352.
8. Ersche KD, Gillan CM, Jones PS, Williams GB, Ward LH, Luijten M, et al. (2016): Carrots and sticks fail to change behavior in cocaine addiction. Science 352:1468–1471.
9. Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hagele C, et al. (2014): Model-based and model-free decisions in alcohol dependence. Neuropsychobiology 70:122–131.
10. Sjoerds Z, de Wit S, van den Brink W, Robbins TW, Beekman AT, Penninx BW, et al. (2013): Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. Transl Psychiatry 3:e337.
11. Tiffany ST, Conklin CA (2000): A cognitive processing model of alcohol craving and compulsive alcohol use. Addiction 95:145–153.
12. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J (2011): Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci 22:490–497.
13. Killcross S, Coutureau E (2003): Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex 13:400–408.
14. de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC (2009): Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J Neurosci 29:11330–11338.
15. O'Doherty JP (2011): Contributions of the ventromedial prefrontal cortex to goal-directed action selection. Ann N Y Acad Sci 1239:118–129.

16. Ostlund SB, Balleine BW (2005): Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. J Neurosci 25:7763–7770.

17. Alexander WH, Brown JW (2011): Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci 14:1338–1344.

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

19. Grüsser SM, Wrase J, Klein S, Hermann D, Smolka MN, Ruf M, et al. (2004): Cue-induced activation of the striatum and medial prefrontal cortex is associated with subsequent relapse in abstinent alcoholics. Psychopharmacology 175:296–302.

20. Beck A, Wustenberg T, Genauck A, Wrase J, Schlagenhauf F, Smolka MN, et al. (2012): Effect of brain structure, brain function, and brain connectivity on relapse in alcohol-dependent patients. Arch Gen Psychiatry 69:842–852.

21. Charlet K, Beck A, Jorde A, Wimmer L, Vollstadt-Klein S, Gallinat J, et al. (2014): Increased neural activity during high working memory load predicts low relapse risk in alcohol dependence. Addict Biol 19:402–414.

22. Barker JM, Torregrossa MM, Taylor JR (2012): Low prefrontal PSA-NCAM confers risk for alcoholism-related behavior. Nat Neurosci 15:1356–1358.

23. Katner SN, Magalong JG, Weiss F (1999): Reinstatement of alcohol-seeking behavior by drug-associated discriminative stimuli after pro-longed extinction in the rat. Neuropsychopharmacology 20:471–479.

24. Flagel SB, Waselus M, Clinton SM, Watson SJ, Akil H (2014): Ante-cedents and consequences of drug abuse in rats selectively bred for high and low response to novelty. Neuropharmacology 76 Pt B: 425–436.

25. Otto AR, Gershman SJ, Markman AB, Daw ND (2013): The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci 24:751–761.

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

27. Brown SA (1985): Expectancies versus background in the prediction of college drinking patterns. J Consult Clin Psychol 53:123–130.

28. Leigh BC (1989): Attitudes and expectancies as predictors of drinking habits: A comparison of three scales. J Stud Alcohol 50:432–440.

29. Reese FL, Chassin L, Molina BS (1994): Alcohol expectancies in early adolescents: Predicting drinking behavior from alcohol expectancies and parental alcoholism. J Stud Alcohol 55:276–284.

30. Goldman MS, Darkes J (2004): Alcohol expectancy multiaxial assessment: A memory network-based approach. Psychol Assess 16:4–15.

31. Deckel AW, Hesselbrock V, Bauer L (1995): Relationship between alcohol-related expectancies and anterior brain functioning in young men at risk for developing alcoholism. Alcohol Clin Exp Res 19: 476–481.

32. Wiers RW, Bartholow BD, van den Wildenberg E, Thush C, Engels RC, Sher KJ, et al. (2007): Automatic and controlled processes and the development of addictive behaviors in adolescents: A review and a model. Pharmacol Biochem Behav 86:263–283.

33. Ide JS, Zhang S, Hu S, Matuskey D, Bednarski SR, Erdman E, et al. (2014): Gray matter volume correlates of global positive alcohol expectancy in non-dependent adult drinkers. Addict Biol 19:895–906.

34. Anderson KG, Schweinsburg A, Paulus MP, Brown SA, Tapert S (2005): Examining personality and alcohol expectancies using func-tional magnetic resonance imaging (fMRI) with adolescents. J Stud Alcohol 66:323–331.

35. Gundersen H, Specht K, Gruner R, Ersland L, Hugdahl K (2008): Separating the effects of alcohol and expectancy on brain activation: An fMRI working memory study. Neuroimage 42:1587–1596.

36. Hogarth L, Field M, Rose AK (2013): Phasic transition from goal-directed to habitual control over drug-seeking produced by conflict-ing reinforcer expectancy. Addict Biol 18:88–97.

37. Ostlund SB, Maidment NT, Balleine BW (2010): Alcohol-paired contextual cues produce an immediate and selective loss of goal-directed action in rats. Front Integr Neurosci 4.

38. Drobes DJ, Carter AC, Goldman MS (2009): Alcohol expectancies and reactivity to alcohol-related and affective cues. Exp Clin Psycho-pharmacol 17:1–9.

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

40. American Psychiatric Association (2000): Diagnostic and Statistical Manual of Mental Disorders, 4th ed, text revision. Washington, DC: American Psychiatric Association.

41. Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

42. Wittchen H-U, Pfister H (1997): DIA-X-Interviews: Manual fur Screening-Verfahren und Interview; Interviewheft Langsschnittunter-suchung (DIA-X-Lifetime); Erganzungsheft (DIA-X-Lifetime); Intervie-wheft Querschnittuntersuchung (DIA-X-12 Monate); Erganzungsheft (DIA-X-12Monate); PC-Programm zur Durchfuhrung des Interviews (Langsund Querschnittuntersuchung); Auswertungsprogramm. Frank-furt, Germany: Swets and Zeitlinger.

43. Jacobi F, Mack S, Gerschler A, Scholl L, Hofler M, Siegert J, et al. (2013): The design and methods of the mental health module in the German Health Interview and Examination Survey for Adults (DEGS1-MH). Int J Methods Psychiatr Res 22:83–99.

44. Garbusow M, Schad DJ, Sommer C, Jünger E, Sebold M, Friedel E, et al. (2014): Pavlovian-to-instrumental transfer in alcohol dependence: A pilot study. Neuropsychobiology 70:111–121.

45. Brainard DH (1997): The Psychophysics Toolbox. Spat Vis 10: 433–436.

46. Sullivan JT, Sykora K, Schneiderman J, Naranjo CA, Sellers EM (1989): Assessment of alcohol withdrawal: The revised clinical institute with-drawal assessment for alcohol scale (CIWA-Ar). Br J Addict 84: 1353–1357.

47. Sobell LC, Sobell MB (1992): Timeline follow-back. A technique for assessing self-reported alcohol consumption. In: Litten RZ, Allen JP, editors. Measuring Alcohol Consumption: Psychosocial and Biochemical Methods. New York: Humana Press, 41–72.

48. Doll BB, Bath KG, Daw ND, Frank MJ (2016): Variability in dopamine genes dissociates model-based and model-free reinforcement learning. J Neurosci 36:1211–1222.

49. Gillan CM, Otto AR, Phelps EA, Daw ND (2015): Model-based learning protects against forming habits. Cogn Affect Behav Neurosci 15: 523–536.

50. Deserno L, Huys QJ, Boehme R, Buchert R, Heinze HJ, Grace AA, et al. (2015): Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proc Natl Acad Sci U S A 112:1595–1600.

51. Deserno L, Wilbertz T, Reiter A, Horstmann A, Neumann J, Villringer A, et al. (2015): Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. Transl Psychiatry 5:e659.

52. Kuntsche E, Knibbe R, Gmel G, Engels R (2006): Replication and validation of the Drinking Motive Questionnaire Revised (DMQ-R, Cooper, 1994) among adolescents in Switzerland. Eur Addict Res 12:161–168.

53. Gmel G, Labhart F, Fallu JS, Kuntsche E (2012): The association be-tween drinking motives and alcohol-related consequences - room for biases and measurement issues? Addiction 107:1580–1589.

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

55. Glascher J, Daw N, Dayan P, O'Doherty JP (2010): States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66: 585–595.

56. Durazzo TC, Tosun D, Buckley S, Gazdzinski S, Mon A, Fryer SL, et al. (2011): Cortical thickness, surface area, and volume of the brain

reward system in alcohol dependence: Relationships to relapse and extended abstinence. Alcohol Clin Exp Res 35:1187–1200.

57. Ashburner J, Friston KJ (2000): Voxel-based morphometry—The methods. Neuroimage 11:805–821.

58. Cooper ML, Frone MR, Russell M, Mudar P (1995): Drinking to regulate positive and negative emotions: A motivational model of alcohol use. J Pers Soc Psychol 69:990–1005.

59. Gowin JL, Ball TM, Wittmann M, Tapert SF, Paulus MP (2015): Individualized relapse prediction: Personality measures and striatal and insular activity during reward-processing robustly predict relapse. Drug Alcohol Depend 152:93–101.

60. Reiter AM, Deserno L, Kallert T, Heinze HJ, Heinz A, Schlagenhauf F (2016): Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. J Neurosci 36:10935–10948.

61. Koya E, Uejima JL, Wihbey KA, Bossert JM, Hope BT, Shaham Y (2009): Role of ventral medial prefrontal cortex in incubation of cocaine craving. Neuropharmacology 56(suppl 1):177–185.

62. Park WK, Bari AA, Jey AR, Anderson SM, Spealman RD, Rowlett JK, et al. (2002): Cocaine administered into the medial prefrontal cortex reinstates cocaine-seeking behavior by increasing AMPA receptor-mediated glutamate transmission in the nucleus accumbens. J Neurosci 22:2916–2925.

63. Childress AR, Mozley PD, McElgin W, Fitzgerald J, Reivich M, O'Brien CP (1999): Limbic activation during cue-induced cocaine craving. Am J Psychiatry 156:11–18.

64. Heinz A, Siessmeier T, Wrase J, Buchholz HG, Grunder G, Kumakura Y, et al. (2005): Correlation of alcohol craving with striatal dopamine synthesis capacity and D2/3 receptor availability: A combined [18F]DOPA and [18F]DMFP PET study in detoxified alcoholic patients. Am J Psychiatry 162:1515–1520.

65. Bossert JM, Stern AL, Theberge FR, Cifani C, Koya E, Hope BT, et al. (2011): Ventral medial prefrontal cortex neuronal ensembles mediate context-induced relapse to heroin. Nat Neurosci 14:420–422.

66. Aas HN, Leigh BC, Anderssen N, Jakobsen R (1998): Two-year longitudinal study of alcohol expectancies and drinking among Norwegian adolescents. Addiction 93:373–384.

67. Duka T, Tasker R, Stephens DN (1998): Alcohol choice and outcome expectancies in social drinkers. Behav Pharmacol 9:643–653.

68. Spanagel R, Holter SM (1999): Long-term alcohol self-administration with repeated alcohol deprivation phases: An animal model of alcoholism? Alcohol Alcohol 34:231–243.

69. Vengeliene V, Bilbao A, Spanagel R (2014): The alcohol deprivation effect model for studying relapse behavior: A comparison between rats and mice. Alcohol 48:313–320.

70. Vengeliene V, Celerier E, Chaskiel L, Penzo F, Spanagel R (2009): Compulsive alcohol drinking in rodents. Addict Biol 14:384–396.

71. Demmel R, Hagen J (2004): The structure of positive alcohol expectancies in alcohol-dependent inpatients. Addict Res Theory 12: 125–140.

72. Zigmond AS, Snaith RP (1983): The hospital anxiety and depression scale. Acta Psychiatr Scand 67:361–370.

73. Mann K, Ackermann K (2000): Die OCDS-G: Psychometrische Kennwerte der deutschen Version der obsessive compulsive drinking scale. Sucht 46:90–100.

74. Lehrl S (2005): Mehrfachwahl-Wortschatz-Intelligenztest MWT-B. Balingen: Spitta.

75. McLeod DR, Griffiths RR, Bigelow GE, Yingling J (1982): An automated version of the digit symbol substitution test (DSST). Behav Res Methods Instrument 14:463–466.

76. Wechsler D (2014): Wechsler Adult Intelligence Scale–fourth edition (WAIS–IV). San Antonio, Texas: Psychological Corporation.