



# Substance use is associated with reduced devaluation sensitivity

Kaileigh A. Byrne<sup>1</sup> · A. Ross Otto<sup>2</sup> · Bo Pang<sup>3</sup> · Christopher J. Patrick<sup>4</sup> · Darrell A. Worthy<sup>3</sup>

Published online: 30 October 2018  
© Psychonomic Society, Inc. 2018

## Abstract

Substance use has been linked to impairments in reward processing and decision-making, yet empirical research on the relationship between substance use and devaluation of reward in humans is limited. We report findings from two studies that tested whether individual differences in substance use behavior predicted reward learning strategies and devaluation sensitivity in a nonclinical sample. Participants in Experiment 1 ( $N = 66$ ) and Experiment 2 ( $N = 91$ ) completed subscales of the Externalizing Spectrum Inventory and then performed a two-stage reinforcement learning task that included a devaluation procedure. Spontaneous eye blink rate was used as an indirect proxy for dopamine functioning. In Experiment 1, correlational analysis revealed a negative relationship between substance use and devaluation sensitivity. In Experiment 2, regression modeling revealed that while spontaneous eyeblink rate moderated the relationship between substance use and reward learning strategies, substance use alone was related to devaluation sensitivity. These results suggest that once reward-action associations are established during reinforcement learning, substance use predicted reduced sensitivity to devaluation independently of variation in eyeblink rate. Thus, substance use is not only related to increased habit formation but also to difficulty disengaging from learned habits. Implications for the role of the dopaminergic system in habitual responding in individuals with substance use problems are discussed.

**Keywords** Substance use · Decision-making · Reward · Devaluation · Habit formation

The dual systems theory of reinforcement learning proposes that learning is mediated through two different neural systems: the model-free system and the model-based system (Daw et al., 2005). The “model-free,” or habit-based, system learns solely from reward prediction errors. In contrast, learning that entails developing a model of the environment with different action-outcome sequence paths in order to maximize long-term future actions reflects the “model-based,” or goal-directed system (Daw et al., 2005; Doya et al., 2002; Friedel et al., 2014). The model-free system is less flexible and entails more automatic processing, while the model-based reinforcement learning system requires more cognitive resources but underlies more flexible, controlled processing (Otto et al., 2013).

**Electronic supplementary material** The online version of this article (<https://doi.org/10.3758/s13415-018-0638-9>) contains supplementary material, which is available to authorized users.

✉ Darrell A. Worthy  
worthyda@tamu.edu

<sup>1</sup> Clemson University, Clemson, SC, USA

<sup>2</sup> McGill University, Montreal, QC, Canada

<sup>3</sup> Texas A&M University, College Station, TX 77843-4235, USA

<sup>4</sup> Florida State University, Tallahassee, FL, USA

The distinction between these two systems has recently received attention from researchers studying counterproductive habitual behaviors such as substance use problems. One theoretical framework for substance abuse proposes that vulnerabilities in decision-making contribute to maladaptive behaviors, such as drug use (Redish, Jensen, & Johnson, 2008). In particular, selective inhibition of the model-based system represents one such vulnerability. Repeated drug use may shift the balance from the model-based to the model-free system, which, in turn, might lead to difficulties in breaking habits (Redish et al., 2008). Dysfunction in the prefrontal cortex may disrupt an individual’s ability to update reward values of non-drug reinforcers. This process might lead to reduced ability to devalue rewards and contribute to compulsory drug use (Goldstein & Volkow, 2011). Indeed, there is evidence that the transition from recreational to addictive substance abuse corresponds to a shift in the use of model-based learning approaches to reliance on model-free strategies (Everitt & Robbins, 2005; Loewenstein & O’Donoghue, 2004; Gillan et al., 2015). For example, alcohol and stimulant addiction has been associated with selective impairment in model-based learning within a two-stage reinforcement learning task (Gillan et al., 2016; Sebold et al., 2014; Voon et al., 2015). Other work using instrumental learning tasks has shown that alcohol-dependent individuals engage in more habit-based

learning than controls (Sjoerds et al., 2013). Similarly, a recent review has proposed that cocaine addiction may impair model-based learning and promote model-free behavior (Lucantonio et al., 2014). However, this relationship has yet to be empirically tested in humans. In addition to substance use problems, impulsive behavior may also contribute to dysregulation in goal-directed and habit-based control (Hogarth et al., 2012). Thus, prior research provides some evidence linking externalizing behaviors with diminished goal-directed learning and reliance on habit-based responding.

While substance abuse is associated with distinct patterns of reward learning, another key factor that may play a role in substance use problems is difficulty in reward disengagement. Prior research indicates that substance abuse is associated with increased perseveration during reversal learning in humans (Ersche et al., 2008; Patzelt et al., 2014). Yet, it is important to note that animal lesion studies have found that distinct neural regions mediate reversal learning and devaluation (Izquierdo & Murray, 2007; Rudebeck & Murray, 2008). Reversal learning entails flexibly updating stimulus-reward associations, but devaluation involves updating reward value information and inhibiting prior learned responses (Rudebeck & Murray, 2008). Beyond variation in learning to respond to action-reward outcomes, substance use problems may correspond to marked declines in the ability to devalue or disengage from those reward-motivated responses. With these issues in mind, the present investigation utilized data from a large representative sample of males and females in the service of two major goals. The first was to examine relations of model-based and model-free components of reinforcement learning with a broad range of substance use behaviors and ascertain whether individual differences in eyeblink rate, an indirect proxy for dopamine functioning might moderate this relationship. The second was to evaluate whether substance use predicts reward disengagement after reinforcement learning has occurred.

One paradigm that has frequently been used to assess reward disengagement is the devaluation procedure, in which reductions in the value attached to a previously rewarded outcome are assessed. Recent work that has employed devaluation procedures in humans demonstrates that goal-directed associative learning predicts heightened devaluation sensitivity. On the other hand, habit-based learning is associated with insensitivity to reward devaluation (Friedel et al., 2014; Gillan et al., 2015). Reward devaluation has been shown to occur in dopamine-free rats (Berridge & Robinson, 1998), indicating that devaluation may not depend on variation in dopaminergic availability. Furthermore, animal research suggests that repeated substance use may lower one's devaluation sensitivity threshold. For example, extended exposure to cocaine in rats has been associated with insensitivity to outcome values and, subsequently, reduced devaluation sensitivity (Leong, Berini, Ghee, & Reichel, 2016; Schoenbaum and Setlow, 2004). It is

plausible, therefore, that substance use may predict diminished devaluation sensitivity in humans as well. However, it is unclear whether striatal dopamine influences devaluation in individuals with substance use tendencies.

Previous research suggests that variation in spontaneous eye blink rate (EBR), a hypothesized index of striatal tonic dopamine functioning (see Jongkees & Colzato, 2016 for a review of EBR methodology and findings), moderates the effects of trait disinhibition and substance abuse on reward wanting and learning, respectively (Byrne, Patrick, & Worthy, 2016a). Specifically, low EBR levels predicted enhanced reward wanting in individuals with high disinhibitory tendencies. In contrast, high EBR predicted better associative learning in individuals who reported high amounts of substance abuse behaviors (Byrne, Patrick, & Worthy, 2016b). Given the specific effects of substance use on associative reward learning, it is reasonable to predict that substance abuse tendencies, but not trait disinhibition, also may be associated with impaired reward disengagement.

To pursue this investigation, we utilized spontaneous EBR, which consistent evidence from several lines of research suggests may provide an indirect physiological index of striatal tonic dopamine functioning (Cavanagh et al., 2014; Elsworth et al., 1991; Groman et al., 2014; Jutkiewicz & Bergman, 2004; Kaminer et al., 2011; Karson, 1983; Taylor et al., 1999), although two recent PET imaging studies failed to observe a positive relationship between dopamine and blink rate (Dang et al., 2017; Sescosse et al., 2017). Prior research, including pharmacological and PET imaging work, demonstrates that faster spontaneous EBR is correlated with elevated dopamine levels in the striatum (Colzato et al., 2008; Karson, 1983; Taylor et al., 1999) and also with dopamine D<sub>2</sub> receptor density in the ventral striatum and caudate nucleus (Dreyer et al., 2010; Groman et al., 2014; Slagter et al., 2015). Moreover, a recent review concluded that in addition to the relationship between EBR and dopaminergic activity, this measure is a reliable predictor of individual differences in performance on tasks that are dopamine-dependent and of aberrations in dopaminergic activity in psychopathology (Jongkees & Colzato, 2016). The Externalizing Spectrum Inventory-Brief Form (Patrick, Kramer, Krueger, & Markon, 2013) was used to assess trait disinhibition and recreational substance abuse tendencies. To test whether substance use tendencies predict both reward learning and disengagement, we used a two-stage reinforcement learning task in combination with a devaluation procedure that has been used in previous research (Gillan et al., 2015).

## Study hypotheses

Given previous work with this task and findings showing that recreational substance-using individuals with high spontaneous

EBRs show enhanced reward learning, we tested the following hypotheses:

H1: Substance use, independent of the EBR index of dopaminergic functioning, will predict habit-based responding on the devaluation procedure, suggesting that the effects of tonic dopamine may not influence goal-directed responding once learning has occurred.

H2: Given previous work reporting that substance-dependent individuals show deficits in model-based learning (Gillan et al., 2016; Sebold et al., 2014; Voon et al., 2015), we hypothesized that substance use would negatively predict model-based strategies during the reinforcement learning phase of the task.

H3: High EBR will positively predict model-based control in the reinforcement learning phase of the task.

H4a: Individuals reporting high levels of recreational substance use with high EBR will show heightened goal-directed responding during the devaluation procedure.

H4b: Individuals reporting high levels of substance use in conjunction with low EBR will show heightened habit-based responding during the devaluation procedure, reflecting a facilitative effect of dopamine on reward learning, but a hindering effect on goal-directed responding.

## Experiment 1

### Method

#### Participants

Sixty-six undergraduate students (47 females, 19 males;  $M_{\text{age}} = 18.42$ ,  $SD_{\text{age}} = 0.71$ ) completed the experiment for partial fulfillment of an introductory psychology course requirement. The study was approved by the Institutional Review Board at Texas A&M University before procedures were implemented.

#### Individual difference measures and experimental task

**spontaneous EBR (Tonic dopamine marker)** Consistent with previous research (Byrne, Norris, & Worthy, 2016a; Byrne, Patrick, & Worthy, 2016b; Chermahini & Hommel, 2010; Colzato, Slagter, van den Wildenberg, & Hommel, 2009; De Jong & Merckelbach, 1990; Ladas, Frantzidis, Bamidis, & Vivas, 2014), we measured spontaneous eye blink using electrooculographic (EOG) recording. We followed the procedure described by Fairclough and Venables (2006) to record EBR: Vertical eye blink activity was collected by attaching Ag-AgCl electrodes above and below the left eye, with a ground electrode placed at the center of the forehead. All EOG signals were filtered at 0.01–10 Hz and amplified by a Biopac

EOG100C differential corneal-retinal potential amplifier. In line with previous research, eye blinks were defined as phasic increases in EOG activity of  $>100 \mu\text{V}$  and less than 500 ms in duration (Byrne, Norris, & Worthy, 2016a; Byrne, Patrick, & Worthy, 2016b; Barbato et al., 2000; Colzato et al., 2009; Colzato et al., 2007). Eye blink frequency was both manually counted and derived using BioPac Acqknowledge software functions, which computed the frequency of amplitude changes of greater than  $100 \mu\text{V}$ , but not duration differences, in order to ensure valid results. The manual and automated EBRs were strongly positively correlated,  $r = 0.97$ ,  $p < 0.001$ . Manual EBR was used for all subsequent statistical analyses.

All recordings were collected during daytime hours of 10 a.m. to 5 p.m., because previous work has shown that diurnal fluctuations in spontaneous EBR can occur in the evening hours (Barbato et al., 2000). A black fixation cross ("X") was displayed at eye level on a wall situated 1 meter from where the participant was seated. Participants were instructed to look in the direction of the fixation cross for the duration of the recording and avoid moving or turning their head. Eye blinks were recorded for 6 minutes under this basic resting condition. Each participant's EBR was quantified by computing the average number of blinks per minute across the 6-min recording interval.

**Externalizing spectrum inventory—brief form** To assess substance use behavior and general externalizing proneness, we administered the Substance Abuse and Disinhibition subscales from the Externalizing Spectrum Inventory – Brief Form (ESI-BF; Patrick et al., 2013). The 18-item Substance Abuse subscale assesses use of and problems with alcohol and drugs, while the 20-item Disinhibition subscale indexes an individual's general proclivity for externalizing problems through questions pertaining to impulsivity, impatient urgency, alienation, irresponsibility, theft, and fraud. Participants responded to the items of each subscale using a 4-point Likert scale (*true*, *somewhat true*, *somewhat false*, and *false*), and responses were coded such that higher scores indicate a greater degree of substance abuse problems and externalizing tendencies, respectively. Prior research has shown strong validity for both subscales in relation to relevant criterion measures (Patrick & Drislane, 2015; Venables & Patrick, 2012). Additionally, within the current study sample, each subscale exhibited high internal consistency (Cronbach's  $\alpha$ s = 0.91 and 0.83 for Substance Abuse and Disinhibition, respectively). Scores for substance use ranged from 0–49 out of a maximum possible score of 72. Figure S1 in the Supplementary Materials displays a histogram showing the frequency distribution of substance use in this sample. Among participants in the sample, 36.92% endorsed a "True" or "Somewhat True" response on at least one marijuana use/problems question, and 40.00% endorsed a "True" or "Somewhat True" response on at least one item indicative of alcohol problems. Table 1

displays the rate of recreational substance use reported for each type of drug assessed.

**Barratt impulsiveness scale** In view of the close relationship between disinhibition and impulsivity, we also administered the 30-item Barratt Impulsiveness Scale (11<sup>th</sup> version; BIS-11). The BIS-11 measures three domains of impulsiveness: motor, cognitive, and nonplanning (Patton et al., 1995). Participants used a 4-point Likert scale from 0–3 (*rarely/never, occasionally, often, and almost always/always*) to indicate the frequency in which they engaged in each questionnaire item. This scale has been shown to have a high degree of internal consistency among college students ( $\alpha = 0.82$ ; Patton et al., 1995).

### Two-stage reinforcement learning devaluation task

**Reinforcement learning phase** A two-stage reinforcement-learning task with a devaluation procedure was used to assess individual differences in devaluation (*unlearning*) of previously valued choices (Gillan et al., 2015; Figure 1). To compare devalued and non-devalued rewards in the devaluation phase, the task included two conditions or “states” (gold and silver) of a two-stage Markov decision process. Each state, which we refer to as a “trial type,” was independent of the other and was denoted using a gold or silver border that clearly outlined choice options. The reward structure and procedure were the same for both trial types, but rewards earned for each trial type were recorded separately. Participants were informed that the points they earned would be stored in a container of the corresponding color.

In the first stage of the task, participants completed two concurrent two-stage reinforcement learning tasks that were structurally equivalent, but had unique stimuli and rewards. Through this procedure, individuals gained experience with two situations (“states”) and reward outcomes in order to base future decisions on (Gläscher et al., 2010). Two options were presented, and participants were given 2.5s to choose one or the other. If a response was not made within this response window, the words “No Response” were displayed on the screen, and the next trial began. For each trial, choosing an option in the first stage cost 5 points, which was displayed on the screen after participants selected an option. Each first-stage option had a 70% probability of leading to a particular state (point box) in the second stage (common state transition) and a 30% probability of leading to a different state in the second stage (rare state transition). The probability that the states in the second stage contained a reward varied across the task based on independent Gaussian random walks ( $SD = 0.025$ ) with a minimum probability of 25% and maximum probability of 75%. Rewards were portrayed in points such that the state-transition point boxes contained either 0 (unrewarded) or 100 (rewarded) points. After the second stage,

**Table 1** Recreational drug use in experiments 1 and 2

	Alcohol	Marijuana	Hallucinogens	Depressants
Experiment 1	40%	37%	6%	5%
Experiment 2	39%	37%	7%	9%

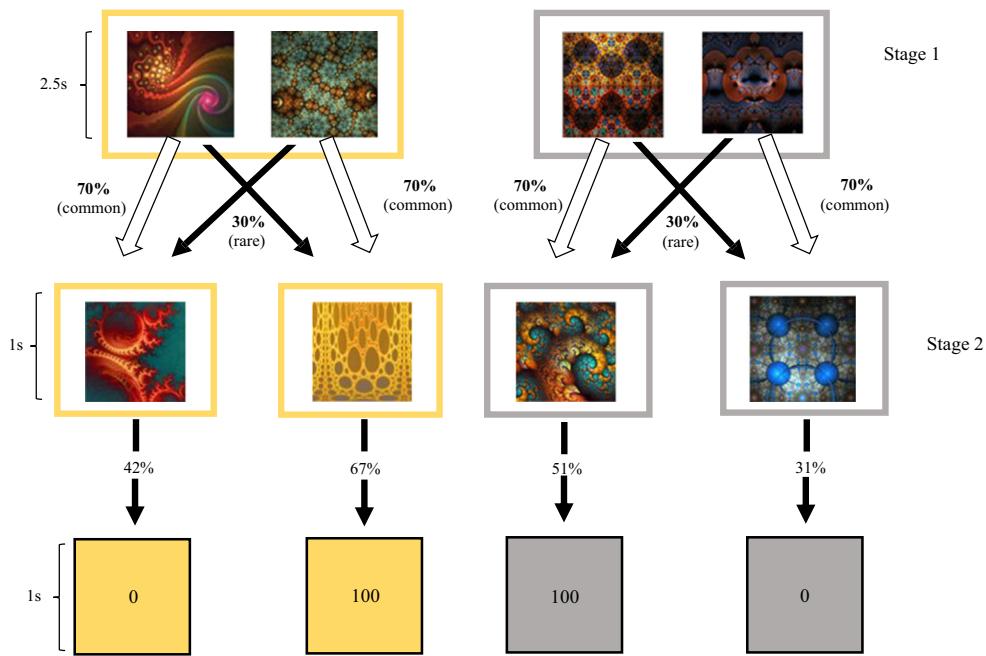
**Note.** Percentages indicate recreational use of each substance reported. Specific hallucinogens included LSD and magic mushrooms and depressants included use of drugs, such as Valium and Xanax for nonmedical reasons.

the points earned were stored in a gold (gold trial type) or silver (silver trial type) container. The cumulative amount earned was displayed throughout the reinforcement learning phase.

Because each trial cost 5 points, if the state in the second stage led to 0 points, then there was a net loss of 5 points for that trial, and the 5 points were deducted from the cumulative total. If the second state yielded 100 points, then there was a net gain of 95 points that was added to the cumulative total. A model-based strategy entails learning which first-stage choice has the common second-stage state transition that is most likely to lead to reward. A model-based learner learns about the chosen first-stage option on common trials, and about the unchosen first-stage option on rare trials. In contrast, a model-free strategy entails staying with a first-stage option if it yielded a reward on the previous trial or switching if the previously selected option was not rewarded; thus, model-free strategies essentially ignore whether the second stage state transition was common or rare.

**Devaluation phase** In the devaluation phase of the task, participants were informed that one of their point containers (i.e., gold) was full and that they would no longer be able to store points in that container. Even if the second stage point box contained 100 points, participants were informed that the points would not be deposited in the container, and they would only be charged the 5 points for that trial. The other point container (i.e., silver) still had room, and participants could still keep points for those trial types. Thus, actions made to earn gold points became devalued.

The outcome of the second stage (either 0 or 100 points) was not displayed in the devaluation phase to ensure that new learning did not contribute to choices in this phase. Participants were informed about this procedural change and advised that the results of their choices (whether or not they received points on each trial) would no longer be shown, but apart from this, nothing about the game had changed and they should continue playing as before. Four trials with no feedback were presented before the devaluation phase, so participants could learn and adjust to this change in feedback presentation before the devaluation procedure began. After these four trials, participants viewed instructions that one of their point containers was completely full. The optimal strategy in this phase was to respond the same way as in the reinforcement learning phase for the valued trial type that still



**Fig. 1** Overview of the reinforcement learning task. Each trial began with one of two trial types (gold or silver). First-stage selections, which cost 5 points, could lead to a common (70%) or rare (30%) second-stage state,

which had a distinct probability of reward that slowed changed based on random Gaussian walks across trials. Selections were not made in the second stage. Figure adapted from Gillan et al., 2015

had room (i.e., silver) and choose not to respond for the devalued trial type where the container was full (i.e., gold).

The two-stage reinforcement learning devaluation task used in this study was the same as that used in previous research (Gillan et al., 2015) with three modifications: (1) we used points rather than monetary rewards (coins) for the second state reward outcome, because participants were recruited from university classes and completed testing for partial course credit rather than being recruited through MTurk and paid for participation as in Gillan et al. (2015); (2) we included 100 trials of the reinforcement learning phase, as opposed to the 200 trials used in some previous studies (e.g., Daw et al., 2011; Gillan et al., 2015), to align with the number of trials used in our own prior work (Byrne, Patrick, & Worthy, 2016a); and (3) we included 50 trials of the devaluation phase rather than 20 trials. The reward structure, stimuli, and design were otherwise identical.

## Procedure

Participants completed the questionnaires and two-stage reinforcement learning devaluation task on PC computers using Psychtoolbox for Matlab (version 2.5). Participants first completed the ESI-BF Substance Abuse and Disinhibition scales along with the BIS-11. Next, participants completed 100 trials of the reinforcement learning phase of the two-stage reinforcement learning devaluation task, followed by 50 trials of the devaluation phase. The instructions for the two-stage reinforcement learning phase and the devaluation phase are presented in the [Supplementary Material](#). The session ended with the 6-min EBR assessment.

## Data analysis

Correlations and multiple hierarchical linear regressions were computed using JASP software ([jasp-stats.org](#)). JASP allows for both frequentist and Bayesian hypothesis tests. For all Bayesian tests we used the default priors from JASP. When testing whether a coefficient from a regression model is significant, we report both the *p* value from a frequentists test, and also the Bayes factor (BF) value for the predictor being tested when entered into a model following the inclusion of all other predictor variables. Bayesian regression in JASP also follows a model comparison approach when evaluating which set of predictors best accounts for a data set. Bayes factors are interpreted as the odds supporting one hypothesis over another. BF values between 3–10 indicate substantial support for the alternative hypothesis over the null hypothesis that a coefficient's true value is zero, BFs between 10–30 indicate strong support, BFs between 30–100 indicate very strong support, and a value above 100 indicates extreme evidence that the alternative hypothesis should be supported over the null (Jeffreys, 1961). In comparative analyses, Wetzels and colleagues generally found that a BF of 3 corresponds to a *p* value of approximately 0.01; many tests that yielded *p* values between 0.01 and 0.05 did not have BFs larger than 3 (Wetzels et al., 2011). Thus, an advantage of including results from Bayesian hypothesis tests is that the conclusions are typically more conservative than those from frequentists tests using an alpha level of 0.05. Additionally, Bayes factors can be interpreted on a continuous scale, unlike *p* values.

Mixed-effects logistic regression analyses were performed using the *lme4* and *brms* modules of the R statistical package,

version 3.0.1. The *lme4* package uses frequentist methods, while the *brms* package uses Bayesian methods to estimate the model parameters. One advantage of using Bayesian methods is that models often fail to converge when using frequentist methods (Burkner, 2017). When reporting results from *lme4* we use *p* values to denote whether regression coefficients differed significantly from zero, and when reporting results from *brms*, we indicate whether the 95% highest density interval from a given coefficient's posterior distribution includes zero. If it does not, then that indicates substantial evidence that the coefficient's true value is not zero.

Trial types (silver coded as 0, gold coded as 1) were computed independently in the analysis such that reward and common/rare states pertained to the previous outcomes for that trial type. For instance, for a trial in which participants selected an option on a gold trial type, the reward and common/rare outcome variables were computed based on the trial preceding that trial type. Reward, second state outcome (common or rare), their interaction, and participants were included as random effects. In line with the analyses of Gillan et al. (2015), we first tested the three-way interaction between Reward, Transition, and Devaluation sensitivity to predict Stay behavior (coded as *stay* 1, *switch* 0). The Reward from the previous trial was coded as *rewarded* 1 and *unrewarded* -1. Transition type on the previous trial was coded as *common* 1 and *rare* 0, and devaluation sensitivity scores were *z*-scored. Devaluation sensitivity was computed as the number of valued trials that participants responded on minus the number of devalued trials that participants responded on. Higher values indicate enhanced devaluation sensitivity.

To compute each participant's model-free and model-based indices, representing the degree to which they engaged in model-free or model-based behavior, we first ran a model predicting Stay from Reward and Transition. The specific syntax for this mixed-effects logistic regression model was: Stay ~ Reward + Transition + Reward:Transition + (1 + Reward + Transition + Reward:Transition | Participant).<sup>1</sup> From this model, individual beta weights were derived. The betas from the Reward variable were used as the model-free metric, and the betas from the Reward X Transition interaction were designated as the model-based metric. Following Gillan et al. (2015), we also regressed Stay behavior on the interaction between Reward, Transition, and Devaluation Sensitivity. The specific syntax for this model is: Stay ~ Reward + Transition + Devaluation + Reward:Transition + Reward:Devaluation + Transition:Devaluation + Reward:Transition:Devaluation + (1 + Reward + Transition + Reward:Transition | Participant).

To evaluate our hypotheses regarding whether the individual difference variables (substance use, EBR, and

disinhibition) would predict performance measures from the task (devaluation sensitivity, model-based index, and model-free index), correlations were first computed among these variables. We predicted that a negative correlation would be observed between substance abuse and the devaluation sensitivity measure. Furthermore, we expected to find a positive correlation between EBR and the model-based index.

Next, to assess relationships between substance use and model-based behavior, we performed regression analyses that (1) tested the prediction that substance use would negatively predict model-based learning, and (2) investigated the possibility that substance use would interact with EBR to influence model-based strategies. These analyses allowed for testing the separate and interactive effects of continuous variations in externalizing tendencies and dopamine levels on reinforcement learning behavior. We employed a hierarchical regression procedure in order to test *a priori* predictions for substance use and its interaction with the EBR proxy of dopaminergic functioning. This approach aligns with analyses in previous work (Byrne, Patrick, & Worthy, 2016a) showing that unique manifestations of externalizing proneness (substance use and disinhibition) exert differential effects on reward processing. The predictors in the model included EBR (striatal dopamine proxy), Substance use, Disinhibition, the EBR X Substance use interaction term, and the EBR X Disinhibition interaction term. In addition, two separate regression analyses were conducted to test the effect of these predictors on devaluation sensitivity and the model-free metric.

Finally, in addition to these hierarchical regression analyses, we also performed mixed effects logistic regression analyses where we included the individual difference variables in the model predicting stay probability from Reward, Transition, and their interaction. This was done to investigate whether devaluation, substance use, disinhibition, or devaluation sensitivity interacted with reward or the Reward X Transition interaction to predict stay behavior. In other words, this model tested whether any of the individual difference variables interacted with the model-free (Reward) or model-based (Reward X Transition) metric. These analyses were exploratory in nature, as we did not have specific predictions regarding individual differences interacting with the model-free or model-based metrics. The specific syntax and results are listed in the [supplementary material](#). All data for both Experiments 1 and 2 are available on the Open Science Framework (<https://osf.io/unk79/>).

## Results

### Correlational analyses

Correlations were used to quantify bivariate relations between the predictor variables (EBR index of striatal dopamine, ESI-BF Substance Abuse, ESI-BF Disinhibition, and BIS-11 impulsiveness) and the criterion measures (devaluation sensitivity index,

<sup>1</sup> Note that *lme4* and *brms* allow the same syntax to be expressed in a simpler form: Stay ~ Reward\*Transition + (1 + Reward\*Transition | Participant). Here we list the extended version for readers unfamiliar with these analysis packages. The \* symbol tells the program to run the model with the interaction term and all the required lower order terms.

model-free index, and the model-based index). Table 2 shows correlations among all variables. Substance Use was significantly negatively correlated with devaluation sensitivity ( $r = -0.26, p < 0.05$ ; Figure 2) and the model-based metric ( $r = -0.30, p < 0.05$ , Figure 3). However, the BF for the substance use—devaluation sensitivity correlation was only 1.26, and the BF for the substance use—model-based metric was 2.65. Thus, these effects are suggestive, but weak. There were no significant correlations for disinhibition, impulsiveness, or the EBR index of striatal dopamine with any of the outcome measures.

### Mixed Effects Logistic Regression Analyses

As described above, we assessed participants' trial-by-trial changes in stay behavior during the reinforcement learning phase of the task (Figure 4) using a model utilized in previous work (Gillan et al., 2015) that included Devaluation sensitivity along with Reward and Transition to predict stay probability on each trial. In line with previous studies (Daw et al., 2011; Gillan et al., 2015), the results demonstrated a significant main effect of Reward ( $\beta = 0.20, p < 0.001$ ), suggesting that participants as a whole engaged in a simple model-free reinforcement strategy during this phase of the experiment. However, the Reward X Transition interaction, indicative of a model-based strategy, was nonsignificant,  $\beta = 0.01, p = 0.82$ . Further results of this analysis are presented in Table S1 of the Supplementary Material.

To examine the individual and interactive effects of externalizing proneness and striatal dopamine on stay behavior, the individual difference variables (EBR index of striatal dopamine, substance use, and disinhibition) were incorporated into an expanded mixed effects logistic regression model. However, while the same effect of reward emerged, as in the simpler model, we found no significant effects of the individual difference factors when predicting stay behavior emerged,  $ps > 0.10$ . There were also no significant interactions between these individual difference variables with Reward ( $ps > 0.10$ ) or Reward X Transition ( $ps > 0.10$ ). The results are reported in Table S2 along with the specific syntax for the model.

### Hierarchical regression analyses

**Devaluation sensitivity** A hierarchical regression analysis was conducted to examine the effect of substance use, disinhibition, and EBR on devaluation sensitivity<sup>2</sup>. In the first step, Gender was added as a covariate, but did not significantly predict devaluation ( $p = 0.91$ ). In the second step, scores for

these three individual variables were entered into the model,  $R^2 = 0.11, F(4, 60) = 1.43, p = 0.23$ . Substance use emerged as a marginally significant predictor ( $\beta = -0.23, p = 0.07, BF = 1.64$ ), whereas effects for disinhibition and EBR were nonsignificant,  $ps > 0.30$ . In the last step of the model, the Substance use X EBR and Disinhibition X EBR interaction terms were added. At this step, none of the predictors ( $ps > 0.40$ ) were significant, nor did the addition of these terms account for a significant proportion of the variance in devaluation sensitivity,  $\Delta R^2 = 0.001, F(2, 58) = 0.02, p = 0.98$ .

**Model-free index** The hierarchical regression analysis with each participant's model-free index as the outcome variable utilized the same predictors as described for the devaluation sensitivity outcome measure. This metric, as well as the model-based index, was dependent on whether participants stayed with or switched options following a rewarded or unrewarded trial and whether the option on the previous trial led to a common or rare box in the second stage. In this model, neither the predictors nor the omnibus predictions were significant in the first and second steps,  $ps > 0.10$ .

**Model-based index** The final set of hierarchical regression models was performed with the model-based index as the outcome variable. In the first step of the model, Gender was added as a covariate, but was nonsignificant,  $p = 0.40$ . In the second step, Substance use ( $\beta = -0.29, p < 0.05, BF = 3.31$ ) negatively predicted model-based behavior, whereas Disinhibition ( $p = 0.91$ ) and EBR ( $p = 0.34$ ) were not significantly associated with the model-based index. However, the omnibus test was not significant at this step,  $R^2 = 0.10, F(4, 59) = 1.88, p = 0.13$ . In the final step with the interaction terms added, the test of  $R^2$  change was nonsignificant,  $\Delta R^2 = 0.02, F(2, 57) = 0.56, p = 0.58$ , and none of the other predictors were significant at this step,  $ps > 0.05$ .

### Discussion

Experiment 1 supports our first hypothesis (H1) and provides preliminary evidence that substance use alone, independent of our indirect proxy for striatal dopaminergic functioning (EBR) and trait disinhibition, is associated with diminished devaluation sensitivity. This result suggests that substance use is associated with reduced reward disengagement. However, it should be noted that the Bayes factors for this effect indicate only weak support for the substance-use devaluation association, possibly due to insufficient power. Moreover, consistent with our second hypothesis (H2), substance use was negatively associated with model-based behavior during the reinforcement learning phase of the task. Because of the relationship between substance use and both model-based behavior and devaluation sensitivity, we also considered the prospect that

<sup>2</sup> We note that the devaluation sensitivity measure evidenced a somewhat binomial distribution. As such, we also performed a logistic regression analysis in which scores of 5 or below were dichotomized as 0, and scores of 6 and above were coded as 1. Significant results were observed for the first step of the regression model, in which substance use ( $\beta = -0.28, p = 0.03$ ) was associated with reduced devaluation sensitivity; however, the results of the omnibus test fell short of significance,  $F(4, 63) = 2.47, p = 0.07$ .

**Table 2** Correlational analyses for Experiment 1

	Substance Use	Disinhibition	Impulsiveness	EBR	Devaluation	MF Index
Disinhibition	0.20					
Impulsiveness	0.09	0.48**				
EBR	-0.02	0.12	-0.13			
Devaluation	-0.26*	-0.15	-0.22	-0.09		
MF Index	-0.20	0.02	0.19	-0.19	0.07	
MB Index	-0.30*	-0.09	-0.03	-0.10	0.07	0.72**

Impulsiveness refers to scores on the BIS-11 Impulsiveness Scale.

\*\*Significance at the  $p < 0.01$  level.

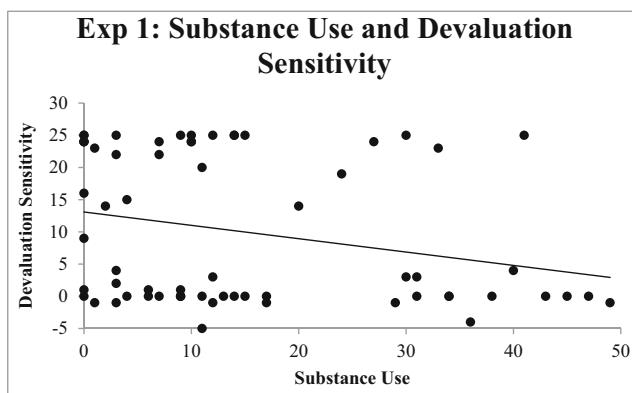
\*Significance at the  $p < 0.05$  level.

an interaction between substance use and model-based behavior may predict devaluation sensitivity. However, as discussed in the [Supplementary Materials](#), this analysis yielded nonsignificant results, which excludes this possibility.

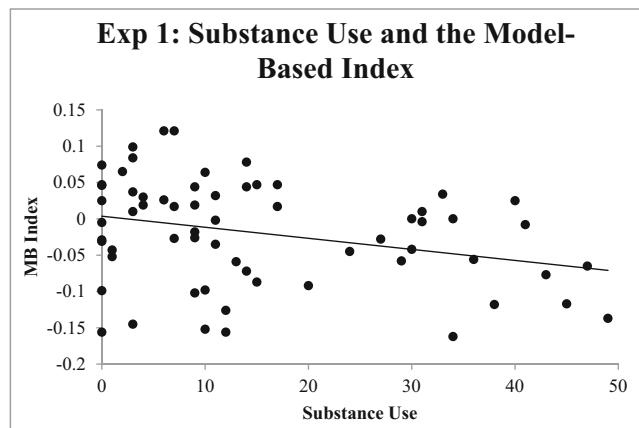
Given the absence of a significant Reward X Transition interaction, and, consequently, the apparent lack of model-based strategy use in overall choice behavior, this finding should be interpreted with caution. One potential limitation to this study that may have contributed to this null finding as well as the observed high correlation between the model-based and model-free indices is that participants performed 100 trials of the task, rather than 200 task trials as employed in previous research by Gillan et al. (2015). Whereas Gillan et al. reported a significant positive relationship between individuals' model-based indices and devaluation sensitivity, our results for Experiment 1 (Table 2) showed a nonsignificant effect ( $r = 0.07, p = 0.58$ ). Thus, 100 trials of the reinforcement learning phase may not have been sufficient for participants to learn which first-stage options had a higher probability of yielding a reward in the second stage. One hundred trials also may have been insufficient to fully capture habit-based responding.

Additionally, a low correlation between substance use and disinhibition was observed. This result was unexpected given the strong association between these measures that has been reported in previous studies (Byrne, Patrick, & Worthy, 2016b; Patrick & Drislane, 2015; Venables & Patrick, 2012). This low correlation can be attributed to a higher proportion of subjects who scored quite low on the disinhibition subscale but quite high on substance abuse subscale. It is conceivable that this experimental sample may have included more students with low disinhibitory tendencies whose higher substance use scores reflected college-age experimentation as opposed to trait-related substance use. A larger sample size may mitigate this incidental issue.

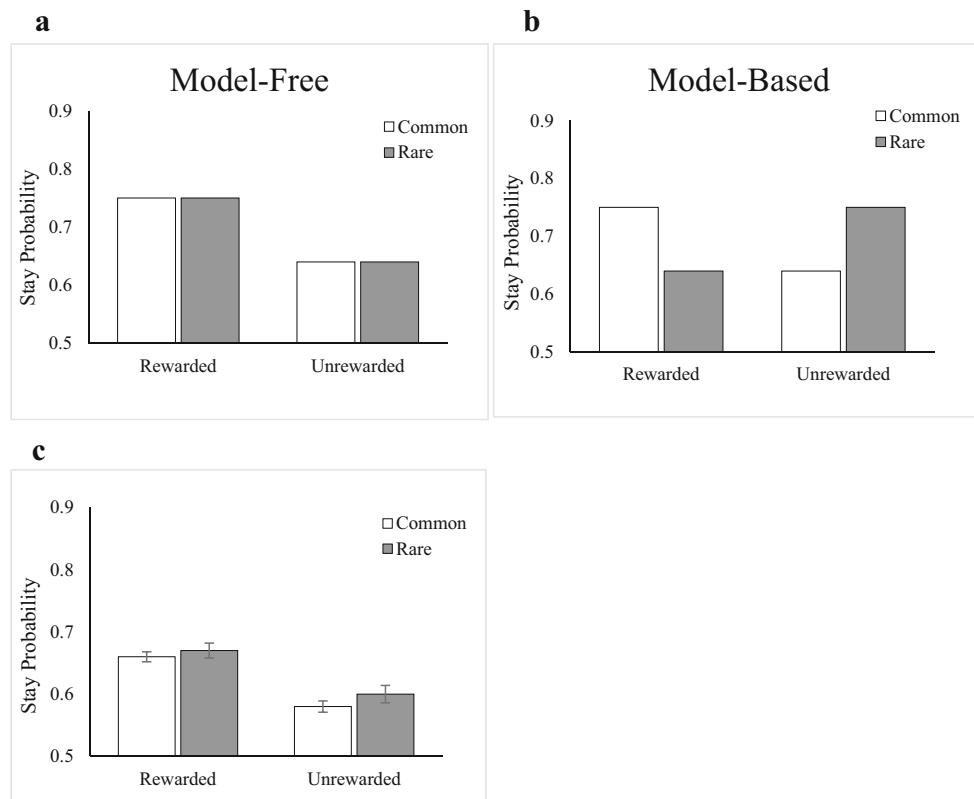
Thus, in Experiment 2 we sought to increase the duration of the reinforcement learning phase to 200 trials in order to test more effectively the relationship between model-based learning and devaluation sensitivity that was observed in Gillan et al. (2015)'s study. Furthermore, we also increase the sample size to provide higher power to test for effects and utilized a sample with representative disinhibition and substance use tendencies. The devaluation phase remained unchanged in Experiment 2.



**Fig. 2** Correlation between substance use and devaluation sensitivity (valued – devalued) in Experiment 1



**Fig. 3** Correlation between substance use and the model-based index in the reinforcement learning task in Experiment 1



**Fig. 4** Analysis of stay behavior in the two-choice reinforcement learning phase of the task in Experiment 1. (a) Depiction of a purely model-free learner whose choices would be predicted by rewarded vs. unrewarded reinforcement strategy. (b) Depiction of a purely model-based learner whose choices would depend on both rewarded vs. unrewarded reinforcement and the common vs. rare transition. (c) Stay proportion

results from the mixed effects logistic regression model averaged across all participants. As reported in the main text, evidence for model-free (main effect of Reward,  $p < 0.001$ ), but not model-based (Reward X Transition interaction) behavior were observed in the 100-trial version of the task in Experiment 1

## Experiment 2

### Method

#### Participants

Ninety-one undergraduate students (61 females, 30 males;  $M_{age} = 18.66$ ,  $SD_{age} = 0.79$ ) completed the experiment for partial fulfillment of a course requirement. The study was approved by the Institutional Review Board at Texas A&M University before procedures were implemented.

#### Procedure

We utilized the same materials and procedure as described in Experiment 1 for Experiment 2, except that participants performed 200 trials of the reinforcement learning phase followed by 50 devaluation phase trials. The longer learning phase was expected to allow individuals to gain more experience learning which second-state transitions are associated with first-stage options and thus have greater opportunity to rely on habit-based responding following the occurrence of learning. We

anticipated that these conditions would be better optimized to yield a clear relationship between model-based learning and devaluation sensitivity, as was found in Gillan et al. (2015). The manual and automated EBRs were again strongly positively correlated,  $r = 0.90$ ,  $p < 0.01$ . In Experiment 2, 36.84% endorsed a “True” or “Somewhat True” response on at least one marijuana use/problems question, and 38.95% endorsed a “True” or “Somewhat True” response on at least one item indicative of alcohol problems (Table 1). Scores for the substance use variable ranged from 0–45 (Figure S2) in this sample.

## Results

### Correlational analyses

Correlations were again used to quantify associations between the predictor variables (EBR index of striatal dopamine, ESI-BF Substance Abuse, ESI-BF Disinhibition, and BIS-11 impulsiveness) and the criterion measures (Devaluation sensitivity index, Model-free index, and the Model-based index). Table 3 displays the results of these analyses. ESI-BF Disinhibition and Substance Abuse subscales were positively

correlated within this sample as expected (Patrick et al., 2013),  $r = 0.49, p < 0.01$ , and each was correlated in turn with BIS-11 impulsiveness scores,  $r_s = 0.58$  and  $0.22$ , respectively,  $p < 0.05$ . As predicted, and consistent with the findings from Experiment 1, substance use was negatively correlated with devaluation sensitivity,  $r = -0.24, p < 0.05$  (Figure 5). However, the BF for this association was only 2.13. Neither disinhibition nor impulsiveness was significantly correlated with any of the outcome measures. However, the EBR index showed a significant negative correlation with devaluation sensitivity ( $r = -0.21, p = 0.04$ ), although the BF for this association was only 1.26.

### Logistic regression analyses

Participants' trial-by-trial alterations in "stay" behavior during the reinforcement learning phase of the task are shown in Figure 6. We again ran the model predicting stay behavior from the interaction between Reward, Transition, and Devaluation sensitivity using both *lme4* and *brms*. Similar to Experiment 1, the regression analysis revealed a significant main effect of Reward ( $\beta = 0.19, p < 0.001$ , 95% HCl: [0.13–0.25]), indicative of a model-free learning strategy. The Reward X Transition interaction coefficient, which quantifies usage of a model-based strategy, was positive but not significant ( $\beta = 0.04, p > 0.05$ , 95% HCl: [-0.01 – 0.07]). There was a significant effect of Devaluation ( $\beta = 0.31, p < 0.001$ , 95% HCl: [0.15–0.48]), indicating that participants with higher devaluation sensitivity were more likely to stay overall. The Reward X Devaluation interaction coefficient also was significant ( $\beta = 0.09, p < 0.01$ , 95% HCl: [0.03–0.15]), which suggests that the effect of Devaluation sensitivity on stay behavior was stronger following rewarded trials than unrewarded trials. Unlike Gillan et al. (2015), we did not observe a significant three-way interaction, perhaps because we did not provide participants with detailed instructions about the task environment, which may have attenuated model-based responding and the association between model-based responding and devaluation sensitivity. The results for the full model are reported in the *Supplementary Material* in Table S3.

We next added the individual difference variables into the mixed effects model predicting stay behavior. These results are reported in supplementary Table S4. We found the same effects for Reward, Devaluation, and their interaction, but no effects or interactions emerged between Reward, Transition, and the individual difference variables.

### Hierarchical regression results

Separate regression analyses were performed for each of the criterion measures: Devaluation sensitivity, the Model-free index, and the Model-based index.

**Devaluation sensitivity** A hierarchical regression analysis was conducted to examine the effect of Substance use, Disinhibition, and EBR (striatal dopamine index) on Devaluation sensitivity.<sup>3</sup> In the first step, Gender was entered as a covariate,  $p = 0.66$ . In the second step, the first-order terms (Substance use, Disinhibition, and EBR) were entered in the model. Omnibus prediction at this step of the model was significant,  $R^2 = 0.11, F(4, 81) = 2.61, p < 0.05$ . Substance use was a significant predictor of devaluation sensitivity ( $\beta = -0.31, p = 0.01, BF = 6.32$ ). Disinhibition also was a significant predictor ( $\beta = 0.25, p < 0.05, BF = 2.54$ ), although the Bayes factor indicated only weak evidence that the coefficient was non-zero (2.54:1 odds). The EBR coefficient was nonsignificant ( $\beta = -0.14, p = 0.18, BF = 1.81$ ). In the last step, the Substance use X EBR and Disinhibition X EBR interaction terms were entered in the model. The inclusion of these terms did not account for a significant proportion of the variance in devaluation sensitivity,  $\Delta R^2 = 0.01, F(2, 79) = 0.49, p = 0.62$ , and none of the individual predictors emerged as significant at this step,  $p > 0.20$ .

**Model-free index** A parallel set of regression analyses was used to test for effects of the predictor variables on model-free behavior. Gender was entered as a covariate in the first step of the model ( $p = 0.57$ ). Substance use, EBR, and Disinhibition were included in the second step. None of these predictors showed a significant association with the model-free index ( $p > 0.08$ ), and the omnibus test for step 2 was nonsignificant,  $R^2 = 0.05, F(4, 80) = 1.07, p = 0.38$ . In the last step of the model, the addition of the interaction terms accounted for a significant proportion of variance in model-free learning,  $\Delta R^2 = 0.08, F(2, 78) = 3.48, p < 0.01$ . Substance use ( $\beta = -1.04, p < 0.01, BF = 10.02$ ) and the Substance use X EBR interaction term ( $\beta = 1.08, p = 0.01, BF = 7.26$ ) were significant predictors of model-free behavior during the reinforcement learning phase. Disinhibition, EBR, and the Disinhibition X EBR interaction were nonsignificant predictors,  $p > 0.10$ .

Figure 7b shows simple regression lines for the effect of substance use scores on model-free choices at (a) the mean for EBR, (b) one standard deviation above the mean EBR, and (c) one standard deviation below the mean for EBR. Substance use and EBR variables were centered prior to creating the centered interaction terms. The simple regression slope coefficients when centered at the mean for EBR ( $\beta = -0.09, p = 0.48$ ) and at one standard deviation below the mean ( $\beta = 0.36, p = 0.09$ ) were not significant, but the simple regression slope coefficient centered at one standard deviation above the mean

<sup>3</sup> As with Experiment 1, we also conducted logistic regression analyses for the data in Experiment 2. For this model, substance use was significantly related to devaluation sensitivity in the first step of the model ( $\beta = -0.26, p = 0.04$ ), although the omnibus test was nonsignificant,  $F(3, 79) = 1.76, p = 0.16$ . The interaction terms were not significant in the second step of the model,  $p > 0.20$ .

**Table 3** Correlational analyses for Experiment 2

	Substance Use	Disinhibition	Impulsiveness	EBR	Devaluation	MF Index
Disinhibition	0.49**					
Impulsiveness	0.22*	0.58**				
EBR	0.11	0.03	0.16			
Devaluation	-0.24*	0.04	0.03	-0.21*		
MF Index	-0.03	0.15	0.12	0.01	0.02	
MB Index	-0.10	0.07	-0.03	-0.03	0.07	0.11

Impulsiveness refers to scores on the BIS-11 Impulsiveness Scale.

\*\*Significance at the  $p < 0.01$  level.

\*Significance at the  $p < 0.05$  level.

for EBR significantly predicted model-free behavior,  $\beta = -0.53$ ,  $p < 0.01$ . At high levels of EBR, individuals with higher substance use tendencies tended to rely less on model-free strategies during the reinforcement learning phase of the task. This result suggests that the degree to which substance use predicted model-free strategy use varied as a function of dopamine such that individuals reporting high levels of substance use with high dopamine showed the least reliance on model-free strategies.

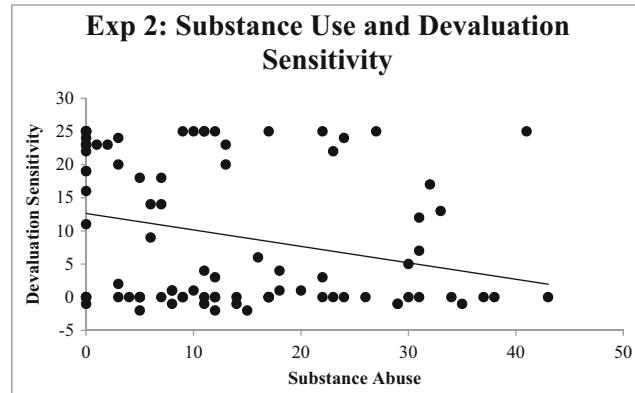
**Model-based index** The same model analysis was conducted for the model-based index as for the devaluation sensitivity and model-free index. The covariate Gender was nonsignificant in the first step,  $p = 0.97$ . All predictors in the second step of the model ( $ps > 0.10$ ) and the omnibus test,  $R^2 = 0.03$ ,  $F(4, 80) = 0.70$ ,  $p = 0.60$ , were nonsignificant. Similarly, in the third step of the model, none of the other predictors ( $ps > 0.30$ ) or the omnibus test,  $R^2 = .04$ ,  $F(6,78) = 0.52$ ,  $p = 0.79$ , were significant.

## Discussion

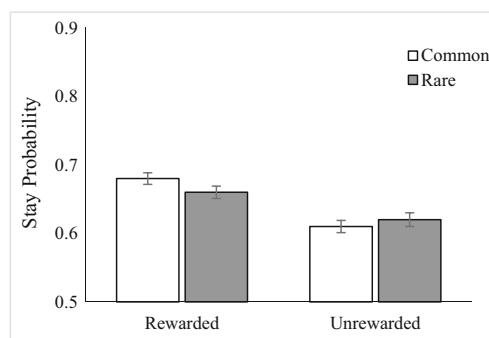
Experiment 2 was designed to better capture model-based behavior. Results showed both a main effect of Reward and a

Reward X Transition interaction were observed, indicating that both model-free and model-based strategies were used in this task, respectively. Consistent with results from Experiment 1, we observed a significant predictive relationship between substance use tendencies and goal-directed behavior in Experiment 2, such that higher substance use was associated with reduced devaluation sensitivity; this result provides support for our second hypothesis. However, while Experiment 1 yielded support for our first hypothesis predicting a negative relationship between substance use and model-based behavior, Experiment 2 did not replicate this effect. Prior work has shown that cognitively demanding model-based strategies yield only negligibly larger rewards in the two-stage task that we used compared with simpler model-free strategies (Akam, Costa, & Dayan, 2015; Kool et al., 2017). Because model-based and model-free computational metrics may not correlate with sizeable differences in task performance and rewards gained, the devaluation procedure, or an alternative two-stage task, may be better suited to characterizing goal-directed versus habit-based learning in future work.

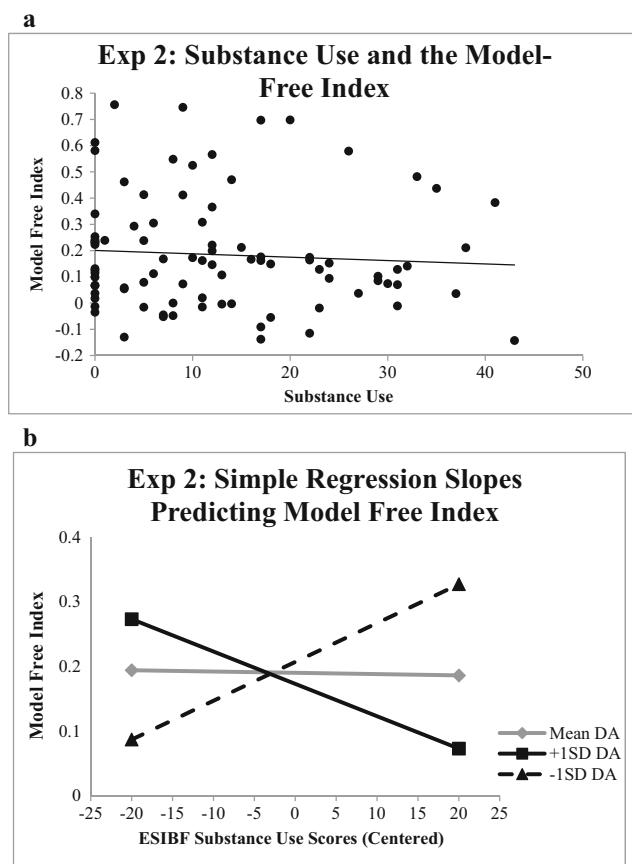
In contrast with the null association for substance use with model-free behavior in Experiment 1, the effect of substance use on model-free behavior in Experiment 2 was moderated



**Fig. 5** Correlation between substance use and devaluation sensitivity (valued – devalued trials) in Experiment 2



**Fig. 6** Stay proportion results from the mixed effects logistic regression model averaged across all participants in Experiment 2. As reported in the main text, evidence for both model-free (main effect of Reward,  $p < 0.01$ ) and model-based (Reward X Transition interaction,  $p = 0.04$ ) behavior were observed in the 200-trial version of the task



**Fig. 7** (a) Correlation between substance use and the model-free index in the reinforcement learning task in Experiment 2. (b) Simple slope regression for the effect of substance use scores (centered at the mean) on the model-free index centered at the mean of EBR, one standard deviation above the mean of EBR, and one standard deviation below the mean of EBR in Experiment 2

by the EBR index of dopamine. One possible explanation for the nonsignificant substance use by EBR interaction in Experiment 1 may be that, due to the limited number of reinforcement learning phase trials and nonsignificant Reward  $\times$  Transition interaction, participants were not able to effectively learn the association between first-stage selections and probability of second-stage outcomes across the learning series. Nevertheless, as we did not observe this result across both experiments, we are cautious in generalizing these findings. Moreover, despite the close association between substance use and trait disinhibition, no association between disinhibition and devaluation sensitivity was observed in either experiment.

## General discussion

Across two experiments, we integrated a reinforcement learning task with a devaluation paradigm to test for associations between substance use tendencies and both reward learning and reward disengagement. Collectively, the results supported

Hypothesis 1, but were not consistently in line with Hypotheses 2–4. Thus, the main finding of study is that substance use is predictive of diminished devaluation sensitivity. Given this evidence for Hypothesis 1, the results fail to support Hypothesis 4a and 4b. The devaluation analyses suggest that once reward-outcome associations are well learned, individuals with substance use problems have difficulty disengaging from habitual responding. These results are consistent with work showing that alcohol dependence predicts increased habit-based learning strategies (Reiter et al., 2016; Sjoerds et al., 2013).

Furthermore, per Hypothesis 2, substance use was expected to show a negative association with model-based behavior. While Experiment 1 yielded provisional support for this prediction, this hypothesis was not supported in Experiment 2. Instead, in Experiment 2, when the EBR index of striatal dopamine was taken into account, specific relationships were found between substance use problems and *model-free*, but not model-based, strategies. Specifically, individuals with substance use problems and high EBR (suggesting heightened levels of striatal tonic dopamine) were less reliant on model-free strategies. These results differ to some extent from previous work suggesting that alcohol and stimulant dependence may be associated with reduced model-based control (Gillan et al., 2016; Sebold et al., 2014; Voon et al., 2015). Conversely, the findings of the current study are in line with previous work that has examined the interactive effects of dopamine and substance use on reward learning (Byrne, Patrick, & Worthy, 2016a).

In Hypothesis 3, we predicted that striatal dopamine (as indexed by EBR) would correlate with model-based control. While previous research has produced evidence for a relationship between increased striatal dopamine levels and model-based behavior (Deserno et al., 2015; Sharp et al., 2015; Wunderlich et al., 2012), findings consistent with this were not found in either of the current experiments. One possible explanation is that prior studies assessed phasic dopaminergic changes, whereas EBR has been shown to indirectly index tonic striatal dopamine (Jongkees & Colzato, 2016). Moreover, recent work shows that genetic variation in prefrontal dopamine regulates model-based learning, while variation in striatal dopamine relates to model-free behavior (Doll et al., 2016). It is therefore plausible that a measure of prefrontal dopamine or phasic dopamine may provide a more sensitive measure of the relationship between dopamine and model-based behavior.

We did not observe a significant relationship between model-based learning and devaluation sensitivity in either Experiment 1 or 2, as was reported in previous work (Gillan et al., 2015). One possible explanation for this difference may be that, unlike the procedure in previous work (Gillan et al., 2015) that relied on monetary compensation during the reinforcement learning phase, our design did not tie monetary

compensation to performance. Additionally, our task in both experiments entailed shorter instructions that did not extensively focus on the structure of the task environment. Thus, it is possible that the lack of compensation and less extensive instructions in our experiments could have dampened the relationship between model-based learning and devaluation sensitivity. We also observed less model-based behavior overall, which could have attenuated any association between devaluation sensitivity and model-based behavior. In addition, our work was conducted at a different time with different participants, than Gillan and colleagues' study. Specifically, our studies were conducted using college-age participants in a campus laboratory setting, whereas Gillan and colleagues' study was conducted on MTurk using participants with a broader age range and vastly different demographics (Paolacci & Chandler, 2014). These differences could have attenuated the link between model-based learning and devaluation sensitivity observed in Gillan et al.'s study.

### Neurobiological mechanisms of substance abuse and habits

The process of reward disengagement from learned stimulus-reward associations has been relatively understudied, especially in humans. While substance use is associated with habit formation (Everitt & Robbins, 2005; Sjoerds et al., 2013), the ability to easily devalue previously rewarding stimuli may reflect a consequence of substance use, or alternatively, may represent a protective factor against addiction formation. Devaluation should prompt reward disengagement, and thus promote "habit-breaking" behavior. Our findings suggest that the effects of substance use entail more than heightened reward wanting and habit-based behavior (Robinson & Berridge, 1993, 2008), resulting in difficulty devaluing rewards and disengaging from learned habits. Furthermore, we suggest that the consequences of problematic substance use may extend beyond dopamine-dependent neural processes.

The question remains that if devaluation is not critically reliant on striatal dopamine, then what are the cognitive and neural mechanisms that account for reduced devaluation sensitivity in individuals with substance use problems? While this issue has yet to be empirically addressed, one possible psychological explanation is that individuals with substance use problems rely on bottom-up signals that usurp cognitive resources needed to engage in top-down control and direct them instead toward reward (Bechara, 2005). Thus, substance use may be associated with a tendency toward using bottom-up reward-based feedback to make decisions, rather than effective model-based strategies and reward devaluation. Two recent studies have demonstrated that compulsive substance use behaviors are perpetuated by a shift from model-based to model-free control that is mediated by reduced activation in the medial prefrontal cortex (Reiter et al., 2016; Sebold et al.,

2017). This decreased activation is associated with deficits in updating information for alternative options during decision-making (Reiter et al., 2016).

Alternatively, current research on the role of the orbitofrontal cortex (OFC) in reward processing also offers some possible answers to the reduced devaluation sensitivity we observed in individuals with substance use problems. Previous neuroimaging research suggests that the OFC computes reward probabilities and codes value representations, regardless of whether the reward is anticipated or received (Kahnt et al., 2010; Li et al., 2016). The OFC thus maintains information about stimuli subjective value and uses that information to guide decisions (Pickens et al., 2003). Furthermore, recent work in primates indicates that OFC volume in area 11/13 is involved in reward devaluation but not reversal learning (Burke et al., 2014). The OFC may therefore play a specific role in devaluation sensitivity by maintaining information about a stimulus' subjective reward value, despite information devaluing that stimulus. Because corticostriatal regions are responsible for working memory updating (Tricomi & Lempert, 2015; Braver & Cohen, 2000; Frank et al., 2001; Frank & O'Reilly, 2006; Cools et al., 2007), it is possible that in individuals with substance use problems, corticostriatal inputs diminish, while OFC activity increases. Thus, updating of stimuli value diminishes in the absence of rewards while prior subjective values of stimuli are maintained, resulting in increased habit responding and reduced devaluation. This explanation is speculative, however, as we did not measure neural activity during this task. Future research is needed to determine the neural mechanism underlying reduced devaluation sensitivity in individuals with substance use problems.

### Limitations

Although the current study findings pertain to individuals with substance use problems, it should be noted that we measured individual differences in substance use in the general population, rather than specifically assessing substance use disorders. It is therefore conceivable that severe or clinical levels of substance use may be associated with different reward learning and devaluation patterns than those observed in our college student sample. We note, however, that in both Experiment 1 and 2 alcohol use and marijuana use were observed at rates of 36–40%, which suggests that our sample encompassed moderate levels of drug and alcohol use. Furthermore, spontaneous EBR is an indirect marker of dopaminergic functioning, and as a result, conclusions should be drawn with caution. Additional measures of dopamine levels, such as positron emission tomography (PET) imaging, are needed to establish a direct relationship between substance use problems and striatal dopamine functioning in reward learning and devaluation contexts. We also note that we did not screen participants for conditions, such as dry eye

syndrome (Tsubota & Nakamori, 1993), and psychological conditions, such as schizophrenia (Karson, Dykman, & Paige, 1990), which have been shown to affect EBR. Finally, we did not control for sleep deprivation or recent use of substances or stimulants, which could alter blink rates, in this study. Future work using EOG to examine substance use and striatal dopamine should examine the influence of these potential factors on blink rate and reward processing.

## Conclusions

Across two experiments, results indicated that after reward contingencies had been established and well learned, substance use alone was associated with reduced devaluation sensitivity, regardless of reinforcement learning strategies or dopaminergic functioning. Moreover, current findings corroborate prior research (Byrne, Patrick, & Worthy, 2016a) demonstrating that substance use tendencies interact with EBR to influence reward learning strategies. Once those action-reward associations are well learned, however, substance use is associated with poorer devaluation of those action-reward contingencies, and thus reduced reward disengagement. These findings demonstrate that problems with substance use extend beyond habit formation; rather, substance use is associated with increased difficulty in disengaging from, or “breaking,” learned habits.

**Acknowledgements** This article was supported by NIA Grant AG043425 to D.A.W.

## References

- Akam, T., Costa, R., & Dayan, P. (2015). Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task. *PLoS Computational Biology*, 11, e1004648.
- Barbato, G., Ficca, G., Muscettola, G., Fichele, M., Beatrice, M., & Rinaldi, F. (2000). Diurnal variation in spontaneous eye-blink rate. *Psychiatry Research*, 93, 145–151.
- Bechara, A. (2005). Decision making, impulse control and loss of will-power to resist drugs: a neurocognitive perspective. *Nature Neuroscience*, 8, 1458–1463.
- Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, 28, 309–369.
- Berridge K. C., Robinson T. E., Aldridge J. W. (2009). Dissecting components of reward: “Liking,” “wanting,” and learning. *Current Opinion in Pharmacology*, 9, 65–73.
- Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. *Control of cognitive processes: Attention and performance XVIII*, 713–737.
- Burke, S. N., Thome, A., Plange, K., Engle, J. R., Trouard, T. P., Gothard, K. M., & Barnes, C. A. (2014). Orbitofrontal cortex volume in area 11/13 predicts reward devaluation, but not reversal learning performance, in young and aged monkeys. *The Journal of Neuroscience*, 34, 9905–9916.
- Burkner, P.C. (2017). Brms: An R package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*. 80, 1–28.
- Byrne, K. A., Norris, D. D., & Worthy, D. A. (2016a). Dopamine, depressive symptoms, and decision-making: the relationship between spontaneous eye blink rate and depressive symptoms predicts Iowa Gambling Task performance. *Cognitive, Affective, & Behavioral Neuroscience*, 16, 23–36.
- Byrne, K. A., Patrick, C. J., & Worthy, D. A. (2016b). Striatal dopamine, externalizing proneness, and substance abuse effects on wanting and learning during reward-based decision making. *Clinical Psychological Science*, 4, 760–774.
- Cavanagh, J. F., Masters, S. E., Bath, K., & Frank, M. J. (2014). Conflict acts as an implicit cost in reinforcement learning. *Nature Communications*, 5, 5394.
- Chermahini, S. A., & Hommel, B. (2010). The (b)link between creativity and dopamine: Spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition*, 115, 458–465.
- Colzato, L. S., Slagter, H. A., van den Wildenberg, W. P., & Hommel, B. (2009). Closing one’s eyes to reality: Evidence for a dopaminergic basis of psychotism from spontaneous eye blink rates. *Personality and Individual Differences*, 46, 377–380.
- Colzato, L. S., van den Wildenberg, W. P., & Hommel, B. (2008). Reduced spontaneous eye blink rates in recreational cocaine users: Evidence for dopaminergic hypoactivity. *PLoS ONE*, 3, e3461.
- Colzato, L. S., van Wouwe, N. C., & Hommel, B. (2007). Spontaneous eye-blink rate predicts the strength of visuomotor binding. *Neuropsychologia*, 45, 2387–2392.
- Cools, R., Sheridan, M., Jacobs, E., & D’Esposito, M. (2007). Impulsive personality predicts dopamine-dependent changes in frontostriatal activity during component processes of working memory. *The Journal of Neuroscience*, 27, 5506–5514.
- Dang, L. C., Samanez-Larkin, G. R., Castrellon, J. J., Perkins, S. F., Cowan, R. L., Newhouse, P. A., & Zald, D. H. (2017). Spontaneous eye blink rate (EBR) is uncorrelated with dopamine D2 receptor availability and unmodulated by dopamine agonism in healthy adults. *eNeuro*, 4, ENEURO-0211.
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans’ choices and striatal prediction errors. *Neuron*, 69, 1204–1215.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8, 1704–1711.
- De Jong, P. J., & Merckelbach, H. (1990). Eyeblink frequency, rehearsal activity, and sympathetic arousal. *International Journal of Neuroscience*, 51, 89–94.
- Deserno, L., Huys, Q. J., Boehme, R., Buchert, R., Heinze, H. J., Grace, A. A., & Schlagenhauf, F. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proceedings of the National Academy of Sciences*, 112, 1595–1600.
- Doll, B. B., Bath, K. G., Daw, N. D., & Frank, M. J. (2016). Variability in dopamine genes dissociates model-based and model-free reinforcement learning. *Journal of Neuroscience*, 36, 1211–1222.
- Doya, K., Samejima, K., Katagiri, K. I., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Computation*, 14, 1347–1369.
- Dreyer, J. K., Herrik, K. F., Berg, R. W., & Hounsgaard, J. D. (2010). Influence of phasic and tonic dopamine release on receptor activation. *The Journal of Neuroscience*, 30, 14273–14283.
- Elsworth, J. D., Lawrence, M. S., Roth, R. H., Taylor, J. R., Mailman, R. B., Nichols, D. E., & Redmond, D. E. (1991). D1 and D2 dopamine receptors independently regulate spontaneous blink rate in the rhesus monkey. *Journal of Pharmacology and Experimental Therapeutics*, 259, 595–600.
- 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.
- Everitt B. J., Robbins T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, 8, 1481–1489.
- Fairclough, S. H., & Venables, L. (2006). Prediction of subjective states from psychophysiology: A multivariate approach. *Biological Psychology*, 71, 100–110.
- Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., & Schlagenauf, F. (2014). Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Frontiers in Human Neuroscience*, 8, 587.
- Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, 1, 137–160.
- Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience*, 120, 497–517.
- Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *Elife*, 5, e11305.
- Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience*, 15, 523–536.
- Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66, 585–595.
- Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews Neuroscience*, 12, 652–669.
- Groman, S. M., James, A. S., Seu, E., Tran, S., Clark, T. A., Harpster, S. N., & Elsworth, J. D. (2014). In the blink of an eye: relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate. *The Journal of Neuroscience*, 34, 14443–14454.
- Hogarth, L., Chase, H. W., & Baess, K. (2012). Impaired goal-directed behavioural control in human impulsivity. *The Quarterly Journal of Experimental Psychology*, 65, 305–316.
- Izquierdo, A., & Murray, E. A. (2007). Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. *The Journal of Neuroscience*, 27, 1054–1062.
- Jeffreys, H. (1961). Theory of probability. Oxford, England: Oxford University Press.
- Jongkees, B. J., & Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience & Biobehavioral Reviews*, 71, 58–82.
- Jutkiewicz, E. M., & Bergman, J. (2004). Effects of dopamine D1 ligands on eye blinking in monkeys: efficacy, antagonism, and D1/D2 interactions. *Journal of Pharmacology and Experimental Therapeutics*, 311, 1008–1015.
- Kahnt, T., Heinze, J., Park, S. Q., & Haynes, J. D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, 107, 6010–6015.
- Kaminer, J., Powers, A. S., Horn, K. G., Hui, C., & Evinger, C. (2011). Characterizing the spontaneous blink generator: an animal model. *Journal of Neuroscience*, 31, 11256–11267.
- Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain*, 106, 643–653.
- Karson, C. N., Dykman, R. A., & Paige, S. R. (1990). Blink rates in schizophrenia. *Schizophrenia Bulletin*, 16, 345–354.
- Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science*, 28, 1321–1333.
- Ladas, A., Frantidis, C., Bamidis, P., & Vivas, A. B. (2014). Eye blink rate as a biological marker of mild cognitive impairment. *International Journal of Psychophysiology*, 93, 12–16.
- Leong, K. C., Berini, C. R., Ghee, S. M., & Reichel, C. M. (2016). Extended cocaine-seeking produces a shift from goal-directed to habitual responding in rats. *Physiology & Behavior*, 164, 330–335.
- Li, Y., Vanni-Mercier, G., Isnard, J., Mauguière, F., & Dreher, J. C. (2016). The neural dynamics of reward value and risk coding in the human orbitofrontal cortex. *Brain*, 139(Pt 4), 1295–1309.
- Loewenstein, G. F. & O'Donoghue, T. (2004). Animal spirits: Affective and deliberative processes in economic behavior. Available at SSRN: <http://ssrn.com/abstract=539843>
- Lucantonio, F., Caprioli, D., & Schoenbaum, G. (2014). Transition from 'model-based' to 'model-free' behavioral control in addiction: involvement of the orbitofrontal cortex and dorsolateral striatum. *Neuropharmacology*, 76, 407–415.
- Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, 24, 751–761.
- Paolacci, G. and Chandler, J. (2014). Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, 23, 184–188.
- Patrick, C. J., & Drislane, L. E. (2015). Triarchic model of psychopathy: Origins, operationalizations, and observed linkages with personality and general psychopathology. *Journal of Personality*, 83, 627–643.
- Patrick, C. J., Kramer, M. D., Krueger, R. F., & Markon, K. E. (2013). Optimizing efficiency of psychopathology assessment through quantitative modeling: Development of a brief form of the Externalizing Spectrum Inventory. *Psychological Assessment*, 25, 1332–1348.
- Patton J. H., Stanford M. S., Barratt E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, 51, 768–774.
- Patzelt, E. H., Kurth-Nelson, Z., Lim, K. O., & MacDonald, A. W. (2014). Excessive state switching underlies reversal learning deficits in cocaine users. *Drug and Alcohol Dependence*, 134, 211–217.
- Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., & Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *The Journal of Neuroscience*, 23, 11078–11084.
- Redish, A. D., Jensen, S., & Johnson, A. (2008). Addiction as vulnerabilities in the decision process. *Behavioral and Brain Sciences*, 31, 461–487.
- Reiter, A. M., Deserno, L., Kallert, T., Heinze, H. J., Heinz, A., & Schlagenauf, F. (2016). Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. *Journal of Neuroscience*, 36, 10935–10948.
- Robbins, T. W. & Everitt, B. J. (1999). Drug addiction: bad habits add up. *Nature*, 398, 567–570.
- Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research Reviews*, 18, 247–291.
- Robinson, T. E., & Berridge, K. C. (2008). The incentive sensitization theory of addiction: some current issues. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363, 3137–3146.
- Rudebeck, P. H., & Murray, E. A. (2008). Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *The Journal of Neuroscience*, 28, 8338–8343.
- Schoenbaum, G., & Setlow, B. (2004). Cocaine makes actions insensitive to outcomes but not extinction: implications for altered orbitofrontal–amygdalar function. *Cerebral Cortex*, 15, 1162–1169.

- Sebold, M., Deserno, L., Nebe, S., Schad, D. J., Garbusow, M., Hägele, C., & Rapp, M. A. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology*, 70, 122–131.
- Sebold, M., Nebe, S., Garbusow, M., Guggenmos, M., Schad, D. J., Beck, A., & Zimmermann, U. S. (2017). When habits are dangerous: alcohol expectancies and habitual decision making predict relapse in alcohol dependence. *Biological Psychiatry*, 82, 847–856.
- Sescousse, G., Ligneul, R., van Holst, R. J., Janssen, L. K., de Boer, F., Janssen, M., & Cools, R. (2017). Spontaneous eye blink rate and dopamine synthesis capacity: Preliminary evidence for an absence of positive correlation. *bioRxiv*, 215178.
- Sharp, M. E., Foerde, K., Daw, N. D., & Shohamy, D. (2015). Dopamine selectively remediates ‘model-based’ reward learning: a computational approach. *Brain*, 139, 355–364.
- Sjoerds, Z., De Wit, S., Van Den Brink, W., Robbins, T. W., Beekman, A. T. F., Penninx, B. W. J. H., & Veltman, D. J. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Translational Psychiatry*, 3, e337.
- Slagter, H. A., Georgopoulou, K., & Frank, M. J. (2015). Spontaneous eye blink rate predicts learning from negative, but not positive, outcomes. *Neuropsychologia*, 71, 126–132.
- Taylor, J. R., Elsworth, J. D., Lawrence, M. S., Sladek, J. R., Jr., Roth, R. H., & Redmond, D. E., Jr. (1999). Spontaneous blink rates correlate with dopamine levels in the caudate nucleus of MPTP-treated monkeys. *Experimental Neurology*, 158, 214–220.
- Tricomi, E., & Lempert, K. M. (2015). Value and probability coding in a feedback-based =learning task utilizing food rewards. *Journal of Neurophysiology*, 113, 4–13.
- Tsubota, K., & Nakamori, K. (1993). Dry eyes and video display terminals. *New England Journal of Medicine*, 328, 584–584.
- Venables N. C., Patrick C. J. (2012). Validity of the Externalizing Spectrum Inventory in a criminal offender sample: Relations with disinhibitory psychopathology, personality, and psychopathic features. *Psychological Assessment*, 24, 88–100.
- Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., & Robbins, T. W. (2015). Disorders of compulsivity: a common bias towards learning habits. *Molecular Psychiatry*, 20, 345–352.
- Wetzels, R., Matzke, D., Lee, M.D., Rouder, J.N., Iverson, G.J., Wagenmakers, E.J. (2011). Statistical evidence in experimental psychology: An empirical Comparison using 855 *t* tests. *Perspective on Psychological Science*, 6, 291–298.
- Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron*, 75, 418–424.