Biological Psychiatry: Celebrating 50 Years

Archival Report

Model-Free and Model-Based Influences in Addiction-Related Behaviors

Stephanie M. Groman, Bart Massi, Samuel R. Mathias, Daeyeol Lee, and Jane R. Taylor

ABSTRACT

BACKGROUND: Disruptions in the decision-making processes that guide action selection are a core feature of many psychiatric disorders, including addiction. Decision making is influenced by the goal-directed and habitual systems that can be computationally characterized using model-based and model-free reinforcement learning algorithms, respectively. Recent evidence suggests an imbalance in the influence of these reinforcement learning systems on behavior in individuals with substance dependence, but it is unknown whether these disruptions are a manifestation of chronic drug use and/or are a preexisting risk factor for addiction.

METHODS: We trained adult male rats on a multistage decision-making task to quantify model-free and model-based processes before and after self-administration of methamphetamine or saline.

RESULTS: Individual differences in model-free, but not model-based, learning prior to any drug use predicted subsequent methamphetamine self-administration; rats with lower model-free behavior took more methamphetamine than rats with higher model-free behavior. This relationship was selective to model-free updating following a rewarded, but not unrewarded, choice. Both model-free and model-based learning were reduced in rats following methamphetamine self-administration, which was due to a decrement in the ability of rats to use unrewarded outcomes appropriately. Moreover, the magnitude of drug-induced disruptions in model-free learning was not correlated with disruptions in model-based behavior, indicating that drug self-administration independently altered both reinforcement learning strategies.

CONCLUSIONS: These findings provide direct evidence that model-free and model-based learning mechanisms are involved in select aspects of addiction vulnerability and pathology, and they provide a unique behavioral platform for conducting systems-level analyses of decision making in preclinical models of mental illness.

Keywords: Computational psychiatry, Dopamine, Drug addiction, Methamphetamine, Model-based reinforcement learning, Model-free reinforcement learning

https://doi.org/10.1016/j.biopsych.2018.12.017

Humans and animals with substance dependence chronically exposed to drugs of abuse have difficulties in making adaptive flexible choices (1–3). A shift in the control of behavior from goal directed to habitual and, ultimately, to the compulsive behavior that has been hypothesized to underlie addiction (4–6) may be the consequence of drug-induced disruptions in the multiple reinforcement learning (RL) systems that guide decision making. Identification of the RL systems that are disrupted by drugs of abuse may be of critical importance for understanding the pathophysiology of addiction, and it requires decision-making tasks that can simultaneously characterize multiple RL strategies within the same individual.

A multistage decision-making (MSDM) task was developed to quantify the degree to which decision making was being influenced by retrospective evaluations of past choices and outcomes (known as model-free RL) and/or by prospective evaluations of future outcomes (known as model-based RL) in healthy individuals (7,8) and in individuals with mental illness (9,10), including addiction (11). It has been argued that the abnormal behaviors in this task observed in individuals with substance dependence reflect a loss of model-based control over behavior (11). However, it also possible that the pattern of decision making observed in individuals with addiction reflects an amplification of the model-free system or even reflects disruptions in both model-free and model-based processes. Nevertheless, an understanding of how these learning systems are affected by chronic drug use has not been fully elucidated. Moreover, the influence of these learning systems on the pathogenesis of drug-taking behaviors is unknown. Inflexible choice behavior that exists prior to any drug exposure could be causally related to components of addiction propensity. Thus, a longitudinal and quantitative analysis of model-free and model-based learning is needed to determine whether the learning systems mediating drug-taking behaviors differ from those modified by drug use.

We have recently developed a translationally inspired variant of the MSDM task for use in rodents (12) that is capable of independent and longitudinal examination of model-free and model-based behavior. Here, we assessed decision making in rats before and after methamphetamine self-administration to

SEE COMMENTARY ON PAGE 886

 ISSN: 0006-3223

determine the precise roles of model-free and model-based processes in the pathophysiology of addiction-like behaviors.

METHODS AND MATERIALS

Subjects

Male Long-Evans rats (N = 60) were obtained from Charles River Laboratories (Raleigh, NC) at approximately 6 weeks of age.

MSDM Task

Rats were trained to make operant-based choices in an MSDM task that was designed to parallel the MSDM task used in humans (Figure 1A). The training protocol has been described in detail elsewhere (12). At the beginning of each trial (state s_A), rats were presented with two spatially distinct lever operandi. Responses on either of the two levers led to one of the two states (s_B or s_C), each corresponding to the illumination of one pair of port apertures (i.e., left lever response → ports 3 and 4; right lever response → ports 1 and 2). Entry into an illuminated port was probabilistically reinforced using an alternating block schedule (Figure 1A, bottom) with the delivery of a single sucrose pellet.

Rats were trained initially on a version of the MSDM task in which choices in the first stage (s_A) deterministically led to the second stage state (referred to as the deterministic MSDM task) (Supplemental Figure S1) to ensure that they understood the structure of the task (see Supplement). Decision making was then assessed on a probabilistic version of the MSDM task where first-stage choices stochastically led to second-stage states (Figure 1A). In the probabilistic MSDM task, choice of one first-stage option (e.g., s_A) predominantly (70%) led to the illumination of the same second-stage state (e.g., s_B or s_C) that was deterministically assigned to that first-stage choice in the deterministic MSDM task (referred to as a common transition). On a limited number of trials (30%), however, first-stage choices led to the illumination of the second-stage state most often associated with the other first-stage choice (referred to as a rare transition). Second-stage choices were reinforced using the same alternating schedule that was used in the deterministic MSDM task (Figure 1A, bottom). Sessions terminated when rats completed 300 trials or 90 minutes had lapsed, whichever occurred first. Rats completed five sessions

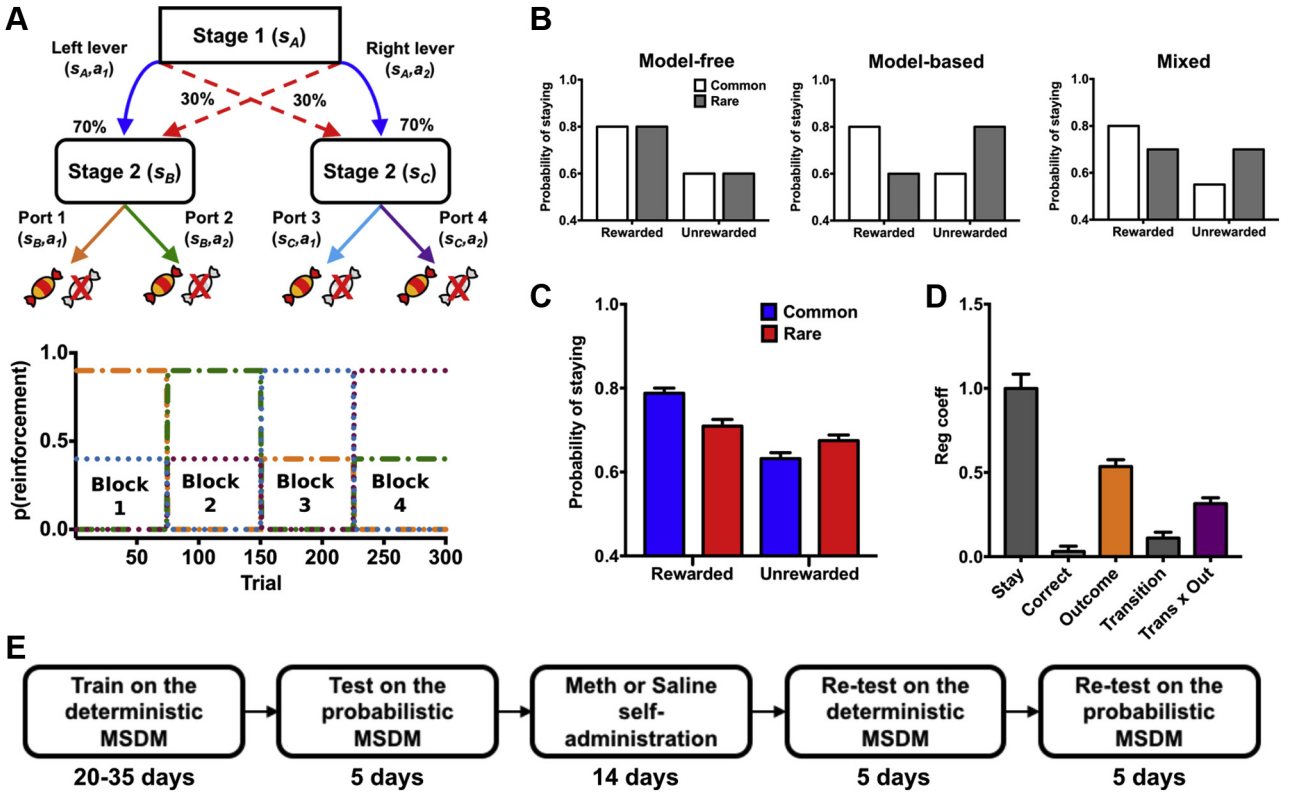

Figure 1. Decision making in the rodent multistage decision-making (MSDM) task. **(A)** Decision making was assessed on a probabilistic MSDM task that paralleled the structure of the human MSDM task (7). **(B)** The probability of staying with the same first-stage choice based on the previous trial outcome (rewarded vs. unrewarded) and the state transition (common transition: open bars; rare transition: gray bars) in theoretical data for a pure model-free agent (left), a pure model-based agent (middle) or an agent using a mix of each strategy (right) in the probabilistic MSDM task. **(C)** The probability of staying with the same first-stage choice based on the previous trial outcome (rewarded vs. unrewarded) and the state transition (common transition: blue bars; rare transition: red bars) in the probabilistic MSDM task. **(D)** The regression weights for the logistic regression model analyzing choice behavior in the probabilistic MSDM task. The weight of the outcome predictor (orange bar) represents the strength of model-free learning, while the transition-by-outcome interaction predictor (purple bar) represents the strength of model-based learning. **(E)** Diagram of the experimental design, with the number of days rats spent in each experimental phase presented below. reg coeff, regression coefficient.

in the probabilistic MSDM task, and trial-by-trial data were collected for the computational analyses described below.

Five days following the self-administration sessions, decision making was reassessed in the deterministic MSDM task for five sessions in a drug-free state to ensure that exposure to methamphetamine did not disrupt understanding of task structure. Decision making was then reassessed in the probabilistic MSDM task for five additional sessions.

Drug Self-administration

Rats (N = 50) were implanted with intrajugular catheters following testing in the MSDM task and were trained to self-administer methamphetamine (0.05 mg/kg/infusion; n = 40) or saline (n = 10) in 6-hour daily sessions for 14 days (Supplement). Rats then underwent forced abstinence for 5 days and were subsequently retested in the MSDM task in a drug-free state.

Data Analysis

Logistic Regression. The trial events that influenced first-stage choices in the MSDM task before and after self-administration were analyzed using a logistic regression model that estimated the likelihood that rats would choose the same first-stage choice based on previous trial events, namely the probability of staying, or p(stay) (Supplement). The model used to analyze choice data in the deterministic MSDM task contained the following predictors: "correct," which coded for first-stage choices that most frequently lead to the highest reinforced second stage; and "outcome," which coded for the outcome of the previous trial. The model used to analyze choice data in the probabilistic MSDM task contained the following predictors: correct; outcome; "transition," which coded for whether the previous state transition was common or rare; and the interaction between transition and outcome. Within this model, the regression coefficient applied to outcome quantifies model-free behavior, and the regression coefficient applied to the transition-by-outcome interaction quantifies model-based behavior.

Choice data in the deterministic MSDM task were also analyzed using a simpler logistic regression model that estimated the likelihood that rats would repeat the same first-stage choice based on whether the previous trial was rewarded or unrewarded [(13); see also Supplement]. This second logistic regression, unlike the first one, permitted an independent analysis of how each trial outcome (rewarded or unrewarded) influenced first-stage choices.

RL Algorithm. Both model-free and model-based learning were also characterized using an RL algorithm (12) that leveraged the strength of the model-based algorithm proposed by Daw *et al.* (7) with a model-free algorithm (14). This model, described in the Supplement, contained a free parameter estimating the weight of model-based learning (β_{MB}) and a different free parameter estimating the weight of model-free learning (β_{MF}).

Characterizing Latent Drug-Taking Phenotypes. The number of drug infusions each rat earned across the 14 days of self-administration were fit with a power function:

$$f(x) = A^{*}x^{B},$$

where x is the session number during the self-administration procedure ($1 \leq x \leq 14$) and A and B are free parameters estimating the scaling factor and rate of growth, respectively. Specifically, the A parameter determines the initial strength of the drug-taking behavior, that is, $A = f(1)$. Rats with higher A values self-administered more methamphetamine initially compared with rats with lower A values (Supplemental Figure S4). The B parameter determines the growth of the function; rats with higher B values increased the number of methamphetamine infusions self-administered at a higher rate than rats with a low B parameter across the self-administration sessions (Supplemental Figure S4).

Statistical Analyses. Statistical analyses were conducted using SPSS (version 25; IBM Corp., Armonk, NY) and R (https://www.R-project.org) and are described in detail in the Supplement.

RESULTS

Assessing Model-Free and Model-Based Learning in Rats

Decision making was assessed in the MSDM task. Data from the deterministic MSDM task are described in Supplemental Figure S2. Choice data in the probabilistic MSDM task (Figure 1A) were examined by calculating the probability that rats would repeat the same first-stage choice according to the outcomes received (rewarded or unrewarded) and the state transition experienced (common or rare) on the immediately preceding trial. According to model-free RL, the probability of repeating the first-stage choice should be influenced only by the previous trial outcome regardless of whether the state transition was common or rare (Figure 1B, left). By contrast, model-based RL posits that the outcome at the second stage should affect the choice of the first-stage option differently based on the state transition that was experienced (Figure 1B, middle). Notably, evidence in humans (7) suggests that individuals use a mixture of model-free and model-based strategies in the MSDM task (Figure 1B, right).

Figure 1C plots the proportion of trials in which rats persisted with the same first-stage choice as a function of the previous trial outcome and state transition. This pattern of results is remarkably similar to that observed in humans (7). The logistic regression (Figure 1D) revealed that the main effect of outcome, which is a measure of model-free learning, was significantly different from zero (t_{59} = 13.14, $p < .001$; orange bars), indicating that rats used second-stage outcomes from the previous trial to guide their first-stage choices on the subsequent trial. The interaction between outcome and transition, which is a measure of model-based learning, was also significantly different from zero (t_{59} = 9.03, $p < .001$; purple bars), indicating that rats took the transition model into account when making their choices. These results demonstrate that rats used both model-free and model-based strategies when making decisions in the MSDM task.

The Roles of Model-Free and Model-Based Learning in Addiction-Related Vulnerability

A diagram of the experimental design is presented in Figure 1E. Across the 14 days of self-administration, the

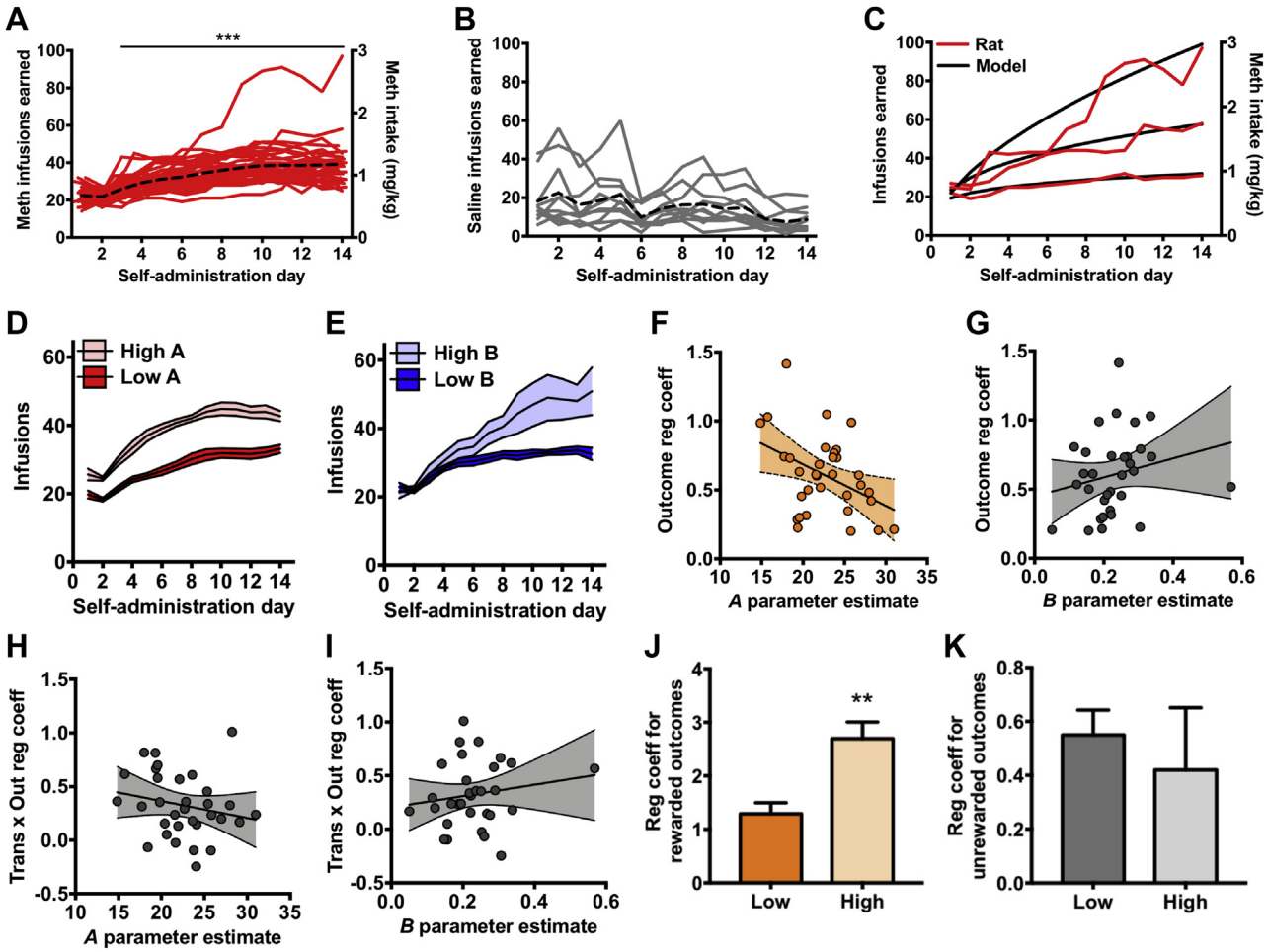

Figure 2. Model-free behavior predicts addiction-relevant behaviors. **(A)** The number of methamphetamine infusions earned in each 6-hour self-administration session across the 14 days for individual rats (red lines) and the average of all rats (black dashed line). ***$p < .001$ compared with the first day of self-administration. See also Supplemental Figure S3. **(B)** The number of saline infusions earned in each 6-hour self-administration across the 14 days for individual rats (gray lines) and the average of all rats (black dashed line). See also Supplemental Figure S3. **(C)** Methamphetamine self-administration data were fit with a power function using maximum likelihood. Drug self-administration data for 3 individual rats are plotted in the red lines. The number of drug infusions predicted by the fitted power function for these 3 rats is represented by the black lines. See also Supplemental Figure S4. **(D)** The number of drug infusions taken across the self-administration sessions in rats with similar rates of escalation in drug use (e.g., *B* parameter) but with low (red; $n = 8$) or high (pink; $n = 8$) values for the initial strength of drug reinforcement (e.g., *A* parameter). See also Supplemental Figure S4. **(E)** The number of drug infusions taken across the self-administration sessions in rats with similar values for the initial strength of drug reinforcement (e.g., *A* parameter) but low (dark blue; $n = 8$) or high (light blue; $n = 8$) rates of escalation in drug use (e.g., *B* parameter). See also Supplemental Figure S4. **(F)** The relationship between model-free behavior (outcome regression coefficient) and the initial strength of drug reinforcement (*A* parameter). **(G)** The relationship between model-free behavior (outcome regression coefficient) and the rate of escalation in drug use (*B* parameter). **(H)** The relationship between model-based behavior (transition-by-outcome regression coefficient) and the initial strength of drug reinforcement (*A* parameter). **(I)** The relationship between model-based behavior (transition-by-outcome regression coefficient) and the rate of escalation in drug use (*B* parameter). **(J)** The regression coefficients from the simple logistic regression indexing the influence of the rewarded outcomes on current choice in rats with low (dark orange) or high (light orange) model-free behavior. **$p < .01$. **(K)** The regression coefficients from the simple logistic regression indexing the influence of the unrewarded outcomes on current choice in rats with low (dark gray) or high (light gray) model-free behavior. meth, methamphetamine; reg coeff, regression coefficient; Trans × Out, transition-by-outcome.

number of methamphetamine infusions earned in each 6-hour session increased ($\chi^2 = 303.94$, $p < .001$) (Figure 2A and Supplemental Figure S3), while the number of saline infusions earned decreased ($\chi^2 = 13.65$, $p < .001$) (Figure 2B and Supplemental Figure S3).

Drug-taking behavior was fit with a power function (Figure 2C) to characterize individual differences in the initial strength of drug reinforcement (*A* parameter) and escalation in drug use (*B* parameter) (Supplemental Figure S4). The power function fit data from individual rats well (SE of the regression = 3.49) and explained nonoverlapping portions of the variance in the total amount of methamphetamine infusions earned when both parameters were included in a multiple linear regression predicting the total infusions earned ($R^2 = .97$; *A*: $\beta = .85$, $p < .001$; *B*: $\beta = .94$, $p < .001$). Therefore, the power function was able to quantify latent drug-taking phenotypes that individually explained unique portions of variance in overall patterns of drug use.

To illustrate this, self-administration data from rats with different values for the initial strength of drug reinforcement [low A values (n = 8): 18.51 ± 0.62; high A values (n = 8): 26.54 ± 0.84; t_{14} = 7.73, $p < .001$] but with similar rates of escalation in drug use [low B values (n = 8): 0.23 ± 0.01; high B values (n = 8): 0.22 ± 0.02; t_{14} = 0.77, p = .46] are presented in Figure 2D. Rats with higher values for the initial strength of drug reinforcement self-administered methamphetamine in higher amounts consistently across the 14 days of self-administration compared with rats with lower values (high vs. low A parameter: χ^2 = 56.74, $p < .001$); however, the rate of escalation in drug use across the self-administration session was similar between the groups (group × time: χ^2 = 0.48, p = .50). Drug self-administration data from rats with different rates of escalation in drug use [low B parameter (n = 8): 0.15 ± 0.001; high B parameter (n = 8): 0.33 ± 0.04; t_{14} = 5.04, $p < .001$] but with similar values for the initial strength of drug reinforcement [low A parameter (n = 8) = 23.40 ± 0.75; high A parameter (n = 8): 21.16 ± 1.08; t_{14} = 1.72, p = .11] are presented in Figure 2E. Initially, rats with higher rates of escalation self-administered similar amounts of methamphetamine compared with rats with lower rates of escalation (high vs. low B parameter: χ^2 = 2.49; p = .12). However, the increase in drug self-administration across sessions was much larger in rats with higher rates of escalation compared with those with lower rates of escalation (group × time: χ^2 = 33.11, $p < .001$). These results indicate that the power function accurately estimates critical components of drug self-administration (6).

Next, we examined the predictive relationship of the model-free and model-based estimates, which were measured prior to any drug exposure, for the drug-taking phenotypes derived from the power function. We found that the outcome coefficient (model-free behavior) accounted for a significant amount of variance in the initial strength of drug reinforcement (R^2 = .17, p = .02) (Figure 2F) but not in the rate of escalation in drug use (R^2 = .04, p = .26) (Figure 2G) (Steiger's Z test comparing the correlation coefficients: Z = 1.21, p = .055). Rats with weaker model-free behavior consistently self-administered more methamphetamine than rats with higher model-free learning across the 14 days of self-administration, but the rate of escalation in drug use was similar between these groups. The transition-by-outcome coefficient (model-based behavior) did not explain a significant amount of variance in either the initial strength of drug reinforcement (R^2 = .05, p = .24) (Figure 2H) or rate of escalation (R^2 = .009, p = .60) (Figure 2I). These data demonstrate a select role of model-free learning in the initiation of drug self-administration, which itself has been found to predict addiction liability in humans (15,16).

To determine whether the relationship between model-free behavior and initial strength of drug reinforcement was due to differences in the influence of specific outcomes on current choice, the regression coefficients from the simpler logistic regression were compared between rats from the lower and upper quartiles of the outcome regression coefficient distribution. The influence of rewarded outcomes was significantly greater in rats with high model-free behavior compared with rats with low model-free behavior (t_{14} = 3.75, p = .002) (Figure 2J); however, the influence of unrewarded outcomes was not different between these groups (t_{14} = 0.53, p = .61) (Figure 2K). Therefore, variation in the initial strength of drug reinforcement may be mediated by preexisting differences in reward-driven, model-free learning.

Drug-Induced Disruptions in Behavior in the Deterministic MSDM Task

The logistic regression coefficients for choice behavior in the deterministic MSDM task were not differently affected in rats that self-administered saline versus those rats that served as a control group for the intrajugular surgeries and operant training (group × time for all predictors: zs < 1.05, ps > .29) (Supplemental Figure S5), so they were combined and collectively served as the comparison group. The p(correct | Stage 1) or p(correct | Stage 2) in the deterministic MSDM task was not differently affected in rats that had self-administered methamphetamine compared with the control group (time × drug: $F_{1,51}$ = 0.75, p = .39) (Figure 3A, B, left). These results indicate that methamphetamine self-administration did not impair the ability of rats to understand the structure of the multistage task and argue against nonselective, drug-induced dysfunctions in operant behavior.

Nevertheless, additional analyses revealed some significant group differences. For example, the probability of staying on the same first-stage choice following an unrewarded trial was differentially altered in rats exposed to methamphetamine compared with control/saline rats (unrewarded group × time: χ^2 = 5.39, p = .02; methamphetamine time: χ^2 = 6.91, p = .009; control/saline time: χ^2 = 0.47, p = .49) (Figure 3A, B, right). A logistic regression analysis indicated that the influence of the previous trial outcome on the current trial choice was reduced in both groups following self-administration (control/saline outcome × time: z = −3.40, $p < .001$; methamphetamine outcome × time: z = −7.43, $p < .001$). This decrement, however, was significantly larger in rats that had self-administered methamphetamine (outcome × time × drug: z = −2.12, p = .02) (Figure 3C, D, left).

The simpler logistic regression model (17) indicated that the likelihood that rats would repeat the same first-stage choice following a rewarded trial was not differentially altered between the experimental groups (drug × time × rewarded: z = 0.74, p = .46). The likelihood that rats would persist with the same first-stage choice following an unrewarded trial, however, was increased following exposure to methamphetamine (drug × time × unrewarded: z = 5.57, $p < .001$; time in methamphetamine group: z = 10.16, $p < .001$) (Figure 3D, right) but not altered in the control/saline group (time in control/saline group: z = 1.03, p = .30) (Figure 3C, right). Therefore, self-administration of methamphetamine selectively disrupted the ability of rats to use negative outcomes appropriately to guide their decision making.

Drug-Induced Disruptions on Model-Free and Model-Based Behavior in the Probabilistic MSDM Task

Logistic regression applied to choice data from the probabilistic MSDM task revealed a significant drug × time × outcome interaction (z = −2.27, p = .02) and a significant drug × time × transition-by-outcome interaction (z = −1.93, p = .05) (Figure 4A, B). These results suggest that both model-free and model-based learning were decreased in rats following

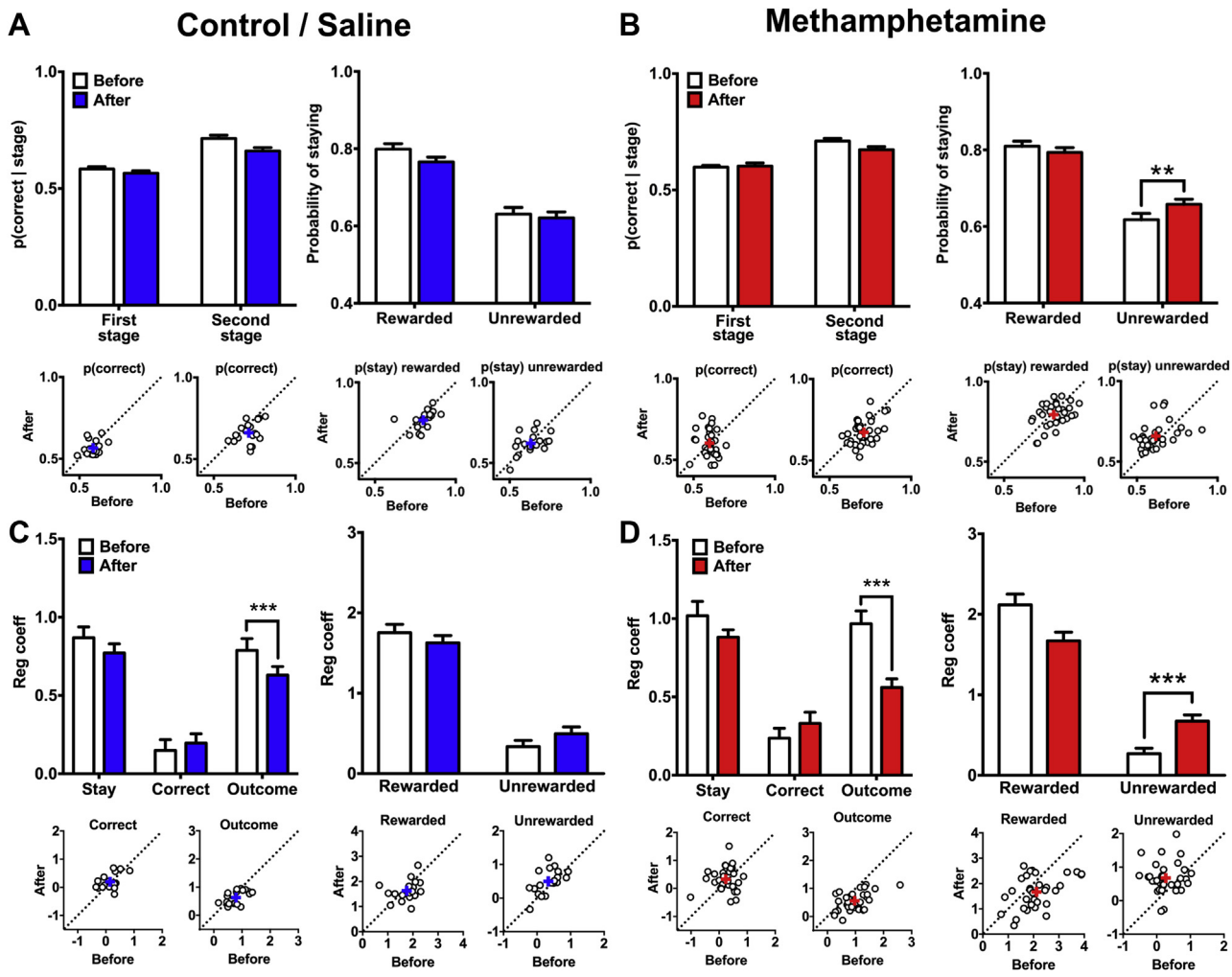

Figure 3. Methamphetamine-induced disruptions in the deterministic MSDM task. **(A)** Left panel: The probability of choosing the first-stage option associated with the highest reinforced second-stage option [*p*(correct | stage 1)] and the probability of choosing the highest reinforced second-stage option [*p*(correct | stage 2)] in the saline/control rats before (open bars) and after (closed bars) self-administration. Right panel: The probability of choosing the same first-stage choice based on the previous trial outcome (i.e., rewarded vs. unrewarded). Below: Scatter plots comparing these dependent measures before and after the self-administration sessions for individual rats are presented below each bar graph, with the mean value represented by the blue symbol. **(B)** Left panel: The probability of choosing the first-stage option associated with the highest reinforced second-stage option [*p*(correct | stage 1)] and the probability of choosing the highest reinforced second-stage option [*p*(correct | stage 2)] in rats before (open bars) and after (closed bars) methamphetamine self-administration. Right panel: The probability of choosing the same first-stage choice based on the previous trial outcome (i.e., rewarded vs. unrewarded). Below: Scatter plots comparing these dependent measures before and after the self-administration sessions for individual rats are presented below each bar graph, with the mean value represented by the red symbol. $^{**}p < .01$. **(C)** The regression coefficient derived from the logistic regression models in control/saline rats before (open bars) and after (closed bars) the self-administration sessions. Left: The regression coefficients from the logistic regression model examining the influence of previous trial outcome on current choice. Right: The regression coefficients from the simple logistic regression model examining the independent influence of rewarded and unrewarded outcomes on current choice. Below: Scatter plots comparing these dependent measures before and after the self-administration sessions for individual rats are presented below each bar graph, with the mean value represented by the blue symbol. $^{***}p < .001$. **(D)** The regression coefficient derived from the logistic regression models in methamphetamine rats before (open bars) and after (closed bars) the self-administration sessions. Left: The regression coefficients from the logistic regression model examining the influence of the previous trial outcome on current choice. Right: The regression coefficients from the simple logistic regression model examining the independent influence of rewarded and unrewarded outcomes on current choice. Below: Scatter plots comparing these dependent measures before and after the self-administration sessions for individual rats are presented below each bar graph, with the mean value represented by the red symbol. $^{***}p < .001$. reg coeff, regression coefficient.

exposure to methamphetamine compared with control/saline rats. Similarly, methamphetamine-induced decrements were detected in the β_{MB} and β_{MF} parameter estimates derived from the hybrid RL model (control/saline Wilcoxon signed rank test: β_{MB}: $p = .81$; β_{MB}: $p = .33$; methamphetamine Wilcoxon signed rank test: β_{MB}: $p = .003$; β_{MF}: $p = .08$) (Figure 4C, D and Supplemental Figure S6). Notably, the balance between model-based and model-free learning measured by calculating the angular coordinate between the outcome and transition-by-outcome regression coefficient, which reflects the relative

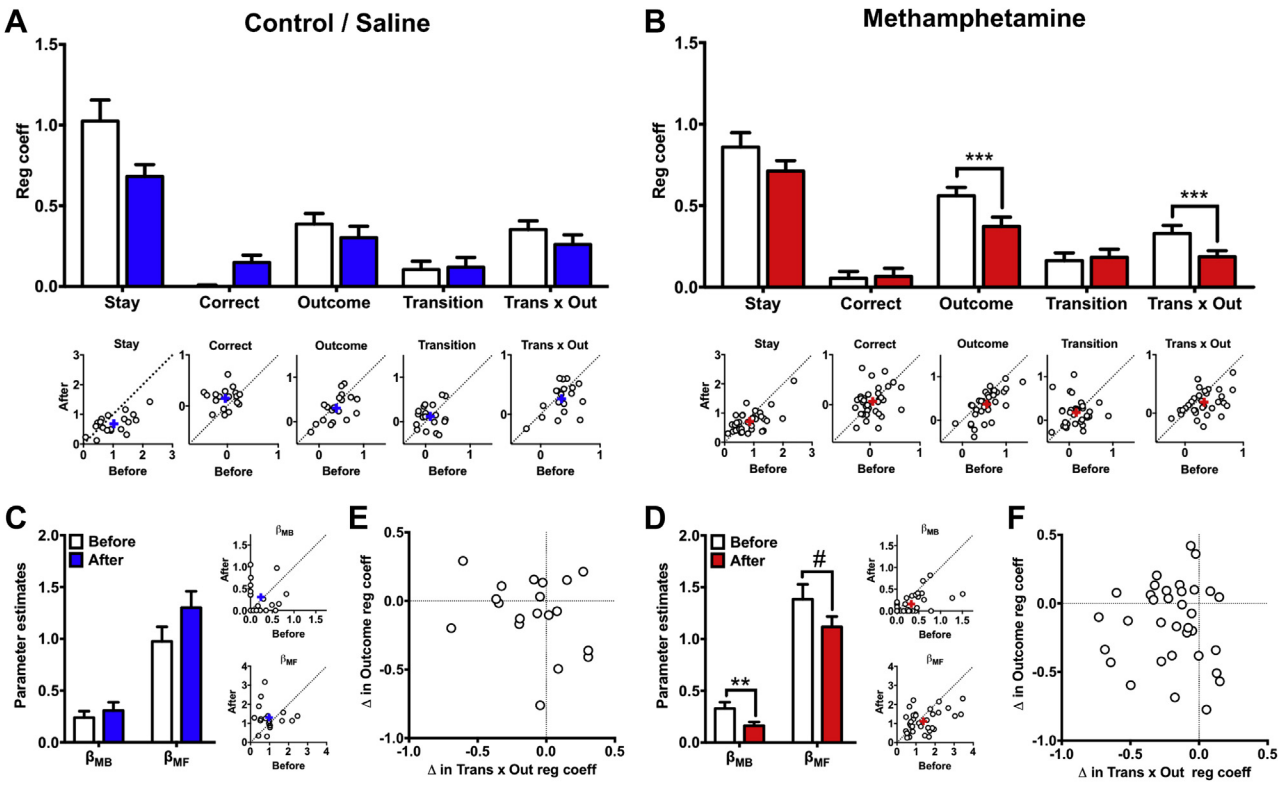

Figure 4. Methamphetamine-induced disruptions in model-free and model-based behavior. **(A)** Regression coefficients from the logistic regression model examining the influence of previous trial events on the likelihood of persisting with the same first-stage choice in the probabilistic multistage decision-making (MSDM) task before (open bars) and after (closed bars) self-administration in control/saline rats. The relationships between regression coefficients before and after self-administration for individual rats are presented below the bar graphs, with the mean value represented by the blue symbol. See also Supplemental Figure S6. **(B)** Regression coefficients from the logistic regression model examining the influence of previous trial events on the likelihood of persisting with the same first-stage choice in the probabilistic MSDM before (open bars) and after (closed bars) self-administration in methamphetamine rats. The relationships between regression coefficients before and after self-administration for individual rats are presented below the bar graphs, with the mean value represented by the red symbol. ***$p < .001$. See also Supplemental Figure S6. **(C)** The free parameter estimates of the weights of model-based learning (β_{MB}) and model-free learning (β_{MF}) obtained from the hybrid reinforcement learning model characterizing choices in the probabilistic MSDM task before (open bars) and after (closed bars) self-administration in control/saline rats. The relationships between these parameters before and after self-administration for individual rats are presented to the right of the bar graphs, with the mean value represented by the blue symbol. See also Supplemental Figure S6. **(D)** The β_{MB} and β_{MF} estimates obtained from the hybrid reinforcement learning model characterizing choices in the probabilistic MSDM task before (open bars) and after (closed bars) self-administration in methamphetamine rats. The relationships between these parameters before and after self-administration for individual rats are presented to the right of the bar graphs, with the mean value represented by the red symbol. **$p < .01$; #$p = .08$. See also Supplemental Figure S6. **(E)** The relationship between the change (before–after self-administration) in the model-free regression coefficient (e.g., outcome) and the change in the model-based regression coefficient (e.g., transition-by-outcome) in control/saline rats. See also Supplemental Figure S7. **(F)** The relationship between the change (before–after self-administration) in the model-free regression coefficient (e.g., outcome) and the change in the model-based regression coefficient (e.g., transition-by-outcome) in methamphetamine rats. See also Supplemental Figure S7. Δ, change; reg coeff, regression coefficient; Trans × Out, transition-by-outcome.

weighting of model-free and model-based strategies, was not affected by methamphetamine self-administration ($p > .53$) (Supplemental Figure S7). This suggests that exposure to methamphetamine did not shift the balance of model-free and model-based influences on behavior but rather disrupted both strategies to a similar degree.

We then examined whether the drug-induced changes in model-free and model-based learning systems were correlated across rats. Difference scores were calculated by subtracting the coefficients measured before self-administration from those measured after self-administration for the outcome and transition-by-outcome coefficients. The difference in model-free behavior was not significantly related to the difference in the model-based behavior in either control/saline rats ($r_s = -.21$, $p = .40$) (Figure 4E) or methamphetamine-exposed rats ($r_s = -.07$, $p = .70$) (Figure 4F). The fact that efficacies of these two learning systems can be adjusted independently suggests that they might not always be in direct competition, as suggested previously (18,19).

DISCUSSION

These data demonstrate that model-free and model-based decision-making processes are involved in distinct aspects of methamphetamine self-administration. Using a novel variant of the MSDM task that parallels that used in humans, we report

that model-free, but not model-based, behavior predicts methamphetamine-taking behaviors in rats. Moreover, the predictive relationship between model-free behavior and subsequent drug self-administration appeared specific to model-free updating following a rewarded trial, suggesting that individuals with low reward-based, model-free updating may be more vulnerable to developing an addiction. Methamphetamine self-administration disrupted the ability of rats to use negative outcomes appropriately, a finding consistent with our previous work using experimenter-administered methamphetamine (20), and consequently decreased both model-free and model-based behaviors. Together, these results indicate that the latent decision-making processes that influence the initiation of drug self-administration differ from those that are affected by chronic drug use and provide a novel framework for disentangling the neurobiological mechanisms involved in addiction vulnerability from those involved in addiction pathology.

Reward-Driven, Model-Free Updating Predicts Addiction-Related Measures

We and others have hypothesized that preexisting disruptions in decision making may be a risk factor for subsequent drug-taking behaviors (21–23). The current study provides direct evidence to support this hypothesis. Model-free behavior predicted methamphetamine-taking behaviors such that rats with weak model-free behavior consistently self-administered more methamphetamine than rats with stronger model-free behavior. By contrast, model-based learning did not predict subsequent drug-taking behaviors, suggesting a unique predictive role of the model-free system in mediating drug reinforcement. Notably, the predictive relationship between model-free behavior and drug self-administration was due to differences in reward-mediated model-free learning, suggesting that dysfunction in reward-based, model-free computations prior to any drug use may be a predictor of initial drug-taking behaviors (24,25).

The negative relationship between initial drug reinforcement and model-free learning observed here may be because rats with disruptions in reward-based, model-free updating are less sensitive to the reinforcing effects of methamphetamine and, consequently, consume greater amounts of drugs. Alternatively, deficits in model-free updating could be associated with heightened drug reinforcement that results in greater drug self-administration. Nevertheless, there is evidence in humans that the initial response to drugs of abuse predicts the likelihood of developing substance dependence later in life (15,16). The *A* parameter—which quantifies initial drug reinforcement—may represent an early addiction vulnerability phenotype. Ongoing studies using the power function to characterize self-administration behavior are examining how variation in initial drug reinforcement (*A* parameter) predicts the development of other addiction-relevant behaviors such as drug reinstatement, extinction, and resistance to punishment to examine this hypothesis.

Model-Free and Model-Based Learning Are Disrupted by Chronic Drug Use

Compulsive drug use has been proposed to be the behavioral manifestation of drug-induced disruptions of the model-based system that lead to a dominance in model-free control over behavior (26). The results described here argue against this hypothesis. Both model-free and model-based behaviors were reduced in rats following methamphetamine self-administration, and the reductions were not due to reductions in motivation, understanding of state transitions, motoric function, differences in operant training, or drug self-administration (Supplement). Moreover, the magnitudes of drug-induced changes in model-free and model-based learning were not related to each other, indicating that these learning systems are independently disrupted by methamphetamine. One potential limitation of the current study is that the differences observed between control and methamphetamine rats was due to the experience of operant training and not specific to the drug. However, this seems unlikely because no differences were observed between the rats that had self-administered saline and received additional operant training and those that served as the surgery control and did not receive additional operant training (Supplemental Figure S5).

The lack of a relationship between drug-induced changes in model-free and model-based learning is important because previous studies in humans have characterized model-free and model-based learning with a single parameter (ω) derived from RL models, which reflects the relative weighting of these two strategies (7). Methamphetamine-dependent individuals have lower ω values compared with healthy controls, which has been interpreted as a loss of model-based behavioral control hypothesized to be the consequence of chronic drug use (11). However, measures that index the relative weighting of model-free and model-based learning can conceal independent differences and/or changes in these RL strategies. Indeed, the relative weighting of model-free and model-based strategies was not disrupted following methamphetamine self-administration despite observing a diminution in both model-free and model-based learning (Supplemental Figure S7). The use of these weighting parameters to characterize these RL systems may, therefore, obfuscate the behavioral processes that are disrupted in mental illness.

Although our MSDM task was developed to emulate the MSDM task used in humans, differences between these paradigms may have contributed to discrepancies in the relative weighting of model-free and model-based strategies between the current study and that reported by Voon *et al.* (11). For example, choice behavior of rats was reinforced with the primary reinforcer food, while humans are reinforced with secondary reinforcers. Another apparent difference is the schedule of reinforcement used in rats compared with humans. In humans, choices in the MSDM task are often reinforced using a schedule that gradually changes across trials through random walk, which may recruit different bio-behavioral mechanisms from those recruited by the alternating block schedule. We have previously reported, however, that model-free and model-based regression coefficients from these two schedules are correlated in drug-naïve rats [Groman *et al.* (12)]. Whether the drug-induced deficits observed here would manifest differently if choices had been reinforced in a random walk schedule is unknown. Alternatively, the lower ω parameter observed in

methamphetamine-dependent individuals might be due to a lack of sensitivity in the analytic approaches used in that study [see supplemental data in Voon *et al.* (11)], which could be revealed using more precise computational analyses such as those reported here.

Dysfunction of these RL systems may be the mechanism by which core features of addiction emerge. For example, the persistence of drug use that occurs regardless of the negative consequences associated with drug use may be due to disruptions in action-value updating following negative feedback, as found in the current study and previously observed in substance-dependent individuals (27–30). In addition, the persistent use of drugs regardless of the positive outcomes associated with abstinence may be due to inaccurate evaluations about future outcomes that bias individuals to choose immediate reinforcement such as drugs. Disruptions in these RL processes may, therefore, be the mechanism by which executive processes are impaired in stimulant-dependent populations.

In summary, the current study demonstrates that model-free RL predicts drug self-administration but that both model-free and model-based RL are impaired following chronic methamphetamine self-administration. Our data suggest that addiction vulnerability may be mediated by preexisting differences in reward-driven, model-free learning, whereas the drug-induced decision-making deficits are due to disruptions in model-based and model-free computations related to unrewarded outcomes. Therefore, the compulsive behaviors observed in individuals with addiction are likely to be the consequence of both preexisting aberrant model-free RL and drug-induced dysfunction of model-free and model-based systems.

ACKNOWLEDGMENTS AND DISCLOSURES

This research was supported by Public Health Service grants from the National Institute on Drug Abuse (Grant Nos. DA041480 and DA043443 [to JRT]), a NARSAD Young Investigator Award from the Brain and Behavior Research Foundation (to SMG), and funding provided by the State of Connecticut.

SMG and JRT were responsible for conceptualization. SMG, DL, and JRT were responsible for methodology. BM and SRM were responsible for software. SMG was responsible for investigation. SMG and JRT were responsible for writing the original draft. SMG, JRT, DL, BM, and SRM were responsible for writing, reviewing, and editing. SMG and JRT were responsible for funding acquisition. All authors approved the final version of this manuscript.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychiatry (SMG, SRM, DL, JRT), Department of Neuroscience (BM, DL), and Department of Psychology (DL, JRT), Yale University, New Haven, Connecticut.

Address correspondence to Stephanie M. Groman, Ph.D., 34 Park Street, New Haven, CT 06515; E-mail: stephanie.groman@yale.edu; or Jane R. Taylor, Ph.D., 34 Park Street, New Haven, CT 06515; E-mail: jane.taylor@yale.edu.

Received Nov 4, 2018; revised Dec 19, 2018; accepted Dec 20, 2018.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2018.12.017.

REFERENCES

1. Jentsch JD, Olausson P, De La Garza R 2nd, Taylor JR (2002): Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. Neuropsychopharmacology 26:183–190.
2. Stalnaker TA, Takahashi Y, Roesch MR, Schoenbaum G (2009): Neural substrates of cognitive inflexibility after chronic cocaine exposure. Neuropharmacology 56:63–72.
3. Ersche KD, Roiser JP, Robbins TW, Sahakian BJ (2008): Chronic cocaine but not chronic amphetamine use is associated with perseverative responding in humans. Psychopharmacology 197:421–431.
4. Dayan P (2009): Dopamine, reinforcement learning, and addiction. Pharmacopsychiatry 42:S56–S65.
5. Jentsch JD, Taylor JR (1999): Impulsivity resulting from frontostriatal dysfunction in drug abuse: Implications for the control of behavior by reward-related stimuli. Psychopharmacology 146:373–390.
6. Everitt BJ, Robbins TW (2005): Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. Nat Neurosci 8:1481–1489.
7. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011): Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.
8. Gläscher 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.
9. Culbreth AJ, Westbrook A, Daw ND, Botvinick M, Barch DM (2016): Reduced model-based decision-making in schizophrenia. J Abnorm Psychol 125:777–787.
10. Sharp ME, Foerde K, Daw ND, Shohamy D (2016): Dopamine selectively remediates "model-based" reward learning: A computational approach. Brain 139:355–364.
11. Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, Enander J, *et al.* (2015): Disorders of compulsivity: A common bias towards learning habits. Mol Psychiatry 20:345–352.
12. Groman SM, Massi B, Mathias SR, Curry DW, Lee D, Taylor JR (2019): Neurochemical and behavioral dissections of decision-making in a rodent multi-stage task. J Neurosci 39:295–306.
13. Akam T, Costa R, Dayan P (2015): Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task. PLoS Comput Biol 11:e1004648.
14. Barraclough DJ, Conroy ML, Lee D (2004): Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 7:404–410.
15. Lambert NM, McLeod M, Schenk S (2006): Subjective responses to initial experience with cocaine: An exploration of the incentive-sensitization theory of drug abuse. Addiction 101:713–725.
16. Haertzen CA, Kocher TR, Miyasato K (1983): Reinforcements from the first drug experience can predict later drug habits and/or addiction: Results with coffee, cigarettes, alcohol, barbiturates, minor and major tranquilizers, stimulants, marijuana, hallucinogens, heroin, opiates and cocaine. Drug Alcohol Depend 11:147–165.
17. Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, *et al.* (2016): Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat Neurosci 19:845–854.
18. Daw ND, Niv Y, Dayan P (2005): Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.
19. Lee SW, Shimojo S, O'Doherty JP (2014): Neural computations underlying arbitration between model-based and model-free learning. Neuron 81:687–699.
20. Groman SM, Rich KM, Smith NJ, Lee D, Taylor JR (2018): Chronic exposure to methamphetamine disrupts reinforcement-based decision making in rats. Neuropsychopharmacology 43:770–780.
21. Redish AD, Jensen S, Johnson A (2008): A unified framework for addiction: Vulnerabilities in the decision process. Behav Brain Sci 31:415–487.

22. Groman SM, Jentsch JD (2013): Identifying the molecular basis of inhibitory control deficits in addictions: Neuroimaging in non-human primates. Curr Opin Neurobiol 23:625–631.
23. Jupp B, Dalley JW (2014): Behavioral endophenotypes of drug addiction: Etiological insights from neuroimaging studies. Neuropharmacology 76(pt B):487–497.
24. Volkow ND, Wang G-J, Fowler JS, Tomasi D, Telang F, Baler R (2010): Addiction: Decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. BioEssays 32:748–755.
25. Huys QJM, Tobler PN, Hasler G, Flagel SB (2014): The role of learning-related dopamine signals in addiction vulnerability. Prog Brain Res 211:31–77.
26. 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(pt B):407–415.
27. Ersche KD, Gillan CM, Jones PS, Williams GB, Ward LHE, Luijten M, *et al.* (2016): Carrots and sticks fail to change behavior in cocaine addiction. Science 352:1468–1471.
28. Parvaz MA, Konova AB, Proudfit GH, Dunning JP, Malaker P, Moeller SJ, *et al.* (2015): Impaired neural response to negative prediction errors in cocaine addiction. J Neurosci 35:1872–1879.
29. Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hägele C, *et al.* (2014): Model-based and model-free decisions in alcohol dependence. Neuropsychobiology 70:122–131.
30. Paulus MP, Hozack N, Frank L, Brown GG, Schuckit MA (2003): Decision making by methamphetamine-dependent subjects is associated with error-rate-independent decrease in prefrontal and parietal activation. Biol Psychiatry 53:65–74.