Published in final edited form as:

Biol Psychiatry. 2019 June 01; 85(11): 936–945. doi:10.1016/j.biopsych.2018.12.017.

Model-free and model-based influences in addiction-related behaviors

Stephanie M. Groman^{*,1}, Bart Massi², Samuel R. Mathias¹, Daeyeol Lee^{1,2,3}, and Jane R. Taylor^{*,1,3}

¹Department of Psychiatry, Yale University

²Department of Neuroscience, Yale University

³Department of Psychology, Yale University

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 substance-dependent individuals, but it is unknown if 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 multi-stage decision-making task to quantifying 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. Model-free and model-based learning were both 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 were not correlated with disruptions in model-based behavior indicating that drug self-administration independently altered both reinforcement-learning strategies.

*Correspondence to be direct to: Stephanie M. Groman, Ph.D. (stephanie.groman@yale.edu); Jane R. Taylor, Ph.D. (jane.taylor@yale.edu), 34 Park Street, New Haven CT 06515.

Author contributions: Conceptualization, S.M.G. and J.R.T. Methodology, S.M.G., D.L., J.R.T. Software, B.M. and S.R.M. Investigation, S.M.G. Writing - Original Draft, S.M.G., J.R.T. Writing, Reviewing & Editing, S.M.G., J.R.T., D.L., B.M., and S.R.M. Funding Acquisition, S.M.G. and J.R.T. All authors approved the final version of this manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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

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 provide a unique behavioral platform for conducting systems-level analyses of decision making in preclinical models of mental illness.

Keywords

drug addiction; model-based reinforcement learning; model-free reinforcement learning; dopamine; methamphetamine; computational psychiatry

Introduction

Substance-dependent humans and animals chronically exposed to drugs of abuse have difficulties making adaptive, flexible choices (1–3). A shift in the control of behavior from goal-directed to habitual and, ultimately, to the compulsive behaviors 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 system(s) that are disrupted by drugs of abuse may be of critical importance for understanding pathophysiology of addiction and requires decision-making tasks that can simultaneously characterize multiple RL strategies within the same individual.

A multi-stage 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 a 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 observed in substance-dependent individuals in this task reflects a loss of model-based control over behavior (11). However, it also possible that the pattern of decision-making observed in addicted individuals reflects an amplification of the model-free system or even 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 in 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 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 (Raleigh, NC) at approximately 6 weeks of age.

Multi-stage decision-making (MSDM) task

Rats were trained to make operant-based choices in a multi-stage decisionmaking (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 and s_C), each corresponding to the illumination of one pair of port apertures (i.e., left lever response $\rightarrow$ ports 3 and 4; right lever response $\rightarrow$ 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; Figure S1) to ensure that they understood the structure of the task (see Supplemental Information). Decision-making was then assessed on a probabilistic version of the MSDM where first stage choices stochastically led to second stage states (Figure 1A). In the probabilistic MSDM, 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 (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 (Figure 1A, bottom). Sessions terminated when rats completed 300 trials or 90 min had lapsed, whichever occurred first. Rats completed five sessions in the probabilistic MSDM and trial-by-trial data collected for the computational analyses described below.

Five days following the self-administration sessions, decision-making was reassessed in the deterministic MSDM 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 re-assessed in the probabilistic MSDM for five additional sessions.

Drug self-administration

Rats (N=50) were implanted with intra-jugular catheters following testing on the MSDM task and trained to self-administer methamphetamine (N=40) or saline (N=10) in 6 h daily sessions for 14 days (Supplemental Information). Rats then underwent forced abstinence for five days and were subsequently re-tested on 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 $p(\text{stay})$ (Supplemental Information). The model used to analyze choice data in the deterministic MSDM 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 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; Supplemental Information). This second logistic regression, unlike the first, permitted an independent analysis of how each trial outcome (rewarded or unrewarded) influenced first-stage choices.

Reinforcement-learning algorithm—Model-free and model-based learning was also characterized using a RL algorithm (12) that leveraged the strength of the model-based algorithm proposed by Daw and colleagues with a model-free algorithm (15). This model, described in Supplemental Information, 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 the rate of growth, respectively. Specifically, the A parameter determines the initial strength of the drug-taking behavior, i.e., $A=f(1)$. Rats with higher A values self-administered more methamphetamine initially compared to rats with lower A values (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 (Figure S4).

Statistical Analyses—Statistical analyses were conducted using SPSS (v. 25; IBM) and R (R-Project; www.R-project.org) and are described in detail in Supplemental Information.

Results

Assessing model-free and model-based learning in rats

Decision-making was assessed in the multi-stage decision-making (MSDM) task. Data from the deterministic MSDM are described in Supplemental Information (Figure S2). Choice data in the probabilistic MSDM (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 only be influenced 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 that 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 ($t(59)=13.14$; $p<0.001$; 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 ($t(59)=9.03$; $p<0.001$; 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 number of methamphetamine infusions earned in each 6 h session increased ($\chi^2=303.94$; $p<0.001$; Figure 2A, S3), while the number of saline infusions earned decreased ($\chi^2=13.65$; $p<0.001$; Figure 2B, 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; Figure S4). The power function fit data from individual rats well (standard error of the regression (S) = 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 ($R^2=0.97$; A : $\beta=0.85$; $p<0.001$; B : $\beta=0.94$; $p<0.001$). 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 value ($N=8$): 18.51 ± 0.62 ; high A value ($N=8$): 26.54

± 0.84 ; $t(14)=7.73$, $p<0.001$) 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 ; $t(14)=0.77$; $p=0.46$) 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 to rats with lower values (high vs. low A parameter: $X^2=56.74$; $p<0.001$); however, the rate of escalation in drug use across the self-administration session was similar between the groups (group $\times$ time: $\chi^2=0.48$; $p=0.50$). 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 ; $t(14)=5.04$, $p<0.001$), 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 ; $t(14)=1.72$; $p=0.11$) are presented in Figure 2E. Initially, rats with higher rates of escalation self-administered similar amounts of methamphetamine compared to rats with lower rates of escalation (high vs. low B parameter: $X^2=2.49$; $p=0.12$). However, the increase in drug self-administration across sessions was much larger in rats with higher rates of escalation compared to those with lower rates of escalation (group $\times$ time: $X^2=33.11$; $p<0.001$). 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, on 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 ($R^2=0.17$; $p=0.02$; Figure 2F) but not in the rate of escalation in drug use ($R^2=0.04$; $p=0.26$; Figure 2G; Steiger's Z test comparing the correlation coefficients: $Z=1.21$; $p=0.055$). 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 ($R^2=0.05$; $p=0.24$; Figure 2H) or rate of escalation ($R^2=0.009$; $p=0.60$; 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 (16, 17).

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 quartile of the outcome regression coefficient distribution. The influence of rewarded outcomes was significantly greater in rats with high model-free behavior compared to rats with low model-free behavior ($t(14)=3.75$; $p=0.002$; Figure 2J); however, the influence of unrewarded outcomes was not different between these groups ($t(14)=0.53$; $p=0.61$; Figure 2K). Therefore, variation in the initial strength of drug reinforcement may be mediated by pre-existing differences in reward-driven, model-free learning.

Drug-induced disruptions in behavior on the deterministic MSDM

The logistic regression coefficients for choice behavior in the deterministic MSDM were not differently affected in rats that self-administered saline versus those rats that served as a control for the intrajugular surgeries and operant training (group $\times$ time for all predictors: z 's < 1.05 ; p 's > 0.29 ; Figure S5), so they were combined and collectively served as the comparison group. The $p(\text{correct} \mid \text{Stage 1})$ or $p(\text{correct} \mid \text{Stage 2})$ in the deterministic MSDM task was not differently affected in rats that had self-administered methamphetamine compared to the control group (time $\times$ drug: $F_{1,51}=0.75$; $p=0.39$; Figure 3A,B, left). These results indicate that methamphetamine self-administration did not impair the ability of rats to understand the structure of the multi-stage task and argue against non-selective, 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 to saline/control rats (Figure 3A,B, right; unrewarded group $\times$ time: $X^2=5.39$; $p=0.02$; methamphetamine: time: $X^2=6.91$; $p=0.009$; saline/control: time: $X^2=0.47$; $p=0.49$). A logistic regression analysis indicated that the influence of the previous trial outcome on the current trial choice was reduced in both groups following selfadministration (control/saline: outcome $\times$ time: $z=-3.40$; $p<0.001$; methamphetamine: outcome $\times$ time: $z=-7.43$; $p<0.001$). This decrement, however, was significantly larger in rats that had self-administered methamphetamine (outcome $\times$ time $\times$ drug: $z=-2.12$; $p=0.02$; Figure 3C,D, left).

The simpler logistic regression model (18) 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 $\times$ time $\times$ rewarded: $z=0.74$; $p=0.46$). The likelihood that rats would persist with the same first stage choice following an unrewarded trial, however, was increased following exposure to methamphetamine (drug $\times$ time $\times$ unrewarded: $z=5.57$; $p<0.001$; time in methamphetamine group: $z=10.16$; $p<0.001$; Figure 3D, right), but not altered in the control/saline group (time in saline/control group: $z=1.03$; $p=0.30$; 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

Logistic regression applied to choice data from the probabilistic MSDM task revealed a significant drug $\times$ time $\times$ outcome interaction ($z=-2.27$; $p=0.02$) and a significant drug $\times$ time $\times$ transition-by-outcome interaction ($z=-1.93$; $p=0.05$; Figures 4A,B). These results suggest that both model-free and model-based learning were decreased in rats following exposure to methamphetamine compared to control/saline rats. Similarly, methamphetamine-induced decrements were detected in the β_{MB} and β_{MF} parameter estimates derived from the hybrid RL model (saline/control: Wilcoxon Signed Rank test: β_{MB} : $p=0.81$; β_{MF} : $p=0.33$; methamphetamine: Wilcoxon Signed Rank test: β_{MB} : $p=0.003$; β_{MF} : $p=0.08$; Figure 4C–D, 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 weighting of model-free and model-based strategies, was not affected by methamphetamine self-administration ($p>0.53$; 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 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 ($r_s=-0.21$; $p=0.40$; Figure 4E) or methamphetamine-exposed rats ($r_s=-0.07$; $p=0.70$; Figure 4F). The fact that efficacies of these two learning systems can be adjusted independently suggests that they might not be always in direct competition, as suggested previously (19, 20).

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 (21), 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 pre-existing disruptions in decisionmaking may be a risk factor for subsequent drug-taking behaviors (22–24). 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 (25, 26).

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 drug. 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 (16, 17). 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 (30). The results described here argue against this hypothesis. Model-free and model-based behaviors were both reduced in rats following methamphetamine self-administration and were not due to reductions in motivation, understanding of state transitions, motoric function, differences in operant training or drug self-administration (Supplemental Information). Moreover, the magnitude 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 (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 to 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 (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. (2015). 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 to 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 biobehavioral 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., 2018). 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 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 irrespective 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 (31–34). 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 biases individuals to choose immediate reinforcement, such as drug. 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 pre-existing 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 addicted individuals are likely to be the consequence of both pre-existing aberrant model-free RL and drug-induced dysfunction of model-free and model-based systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

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

References:

1. Jentsch JD, Olsson P, De La Garza R 2nd, Taylor JR (2002): Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. *Neuropsychopharmacology*, 2002/01/16 26: 183–190. [PubMed: 11790514]
2. Stalnaker TA, Takahashi Y, Roesch MR, Schoenbaum G (2009): Neural substrates of cognitive inflexibility after chronic cocaine exposure. *Neuropharmacology*. 56: 63–72. [PubMed: 18692512]
3. Ersche KD, Roiser JP, Robbins TW, Sahakian BJ (2008): Chronic cocaine but not chronic amphetamine use is associated with perseverative responding in humans. *Psychopharmacol*, 2008/01/25 197: 421–431.
4. Dayan P (2009): Dopamine, Reinforcement Learning, and Addiction. *Pharmacopsychiatry*. 42: S56–S65. [PubMed: 19434556]
5. Jentsch JD, Taylor JR (1999): Impulsivity resulting from frontostriatal dysfunction in drug abuse: implications for the control of behavior by reward-related stimuli. *Psychopharmacol*, 1999/11/07 146: 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. [PubMed: 16251991]
7. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011): Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 2011/03/26 69: 1204–1215. [PubMed: 21435563]
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. [PubMed: 20510862]
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. [PubMed: 27175984]
10. Sharp ME, Foerde K, Daw ND, Shohamy D (2016): Dopamine selectively remediates “model-based” reward learning: a computational approach. *Brain*. 139: 355–64. [PubMed: 26685155]
11. Voon V, Derbyshire K, Ruck C, Irvine MA, Worbe Y, Enander J, et al. (2015): Disorders of compulsivity: a common bias towards learning habits. *Mol Psychiatry*. 20: 345–352. [PubMed: 24840709]
12. Groman SM, Massi B, Mathias SR, Curry DW, Lee D, Taylor JR (2018): Neurochemical and behavioral dissections of decision-making in a rodent multistage task. *J Neurosci*. 2219–18.
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. [PubMed: 26657806]
14. Groman SM, Rich KM, Smith NJ, Lee D, Taylor JR (2018): Chronic Exposure to Methamphetamine Disrupts Reinforcement-Based Decision Making in Rats. *Neuropsychopharmacology*. 43. doi: 10.1038/npp.2017.159.
15. Barraclough DJ, Conroy ML, Lee D (2004): Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci*, 2004/03/09 7: 404–410. [PubMed: 15004564]
16. 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. [PubMed: 16669905]
17. 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: 147–65. [PubMed: 6134605]
18. 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. [PubMed: 27110917]
19. Daw ND, Niv Y, Dayan P (2005): Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*, 2005/11/16 8: 1704–1711. [PubMed: 16286932]

20. Lee SW, Shimojo S, O'Doherty JP (2014): Neural computations underlying arbitration between model-based and model-free learning. *Neuron*. 81: 687–99. [PubMed: 24507199]
21. 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. [PubMed: 28741627]
22. Redish AD, Jensen S, Johnson A (2008): A unified framework for addiction: vulnerabilities in the decision process. *Behav Brain Sci*. 31: 415–487. [PubMed: 18662461]
23. Groman SM, Jentsch JD (2013): Identifying the molecular basis of inhibitory control deficits in addictions: Neuroimaging in non-human primates. *Curr Opin Neurobiol*. 23. doi: 10.1016/j.conb.2013.03.001.
24. Jupp B, Dalley JW (2014): Behavioral endophenotypes of drug addiction: Etiological insights from neuroimaging studies. *Neuropharmacology*. 76 Pt B: 487–497. [PubMed: 23756169]
25. 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. [PubMed: 20730946]
26. Huys QJM, Tobler PN, Hasler G, Flagel SB (2014): The role of learning-related dopamine signals in addiction vulnerability 3. *Prog Brain Res*. 211. doi: 10.1016/B978-0-444-63425-2.00003-9.
27. Park SQ, Kahnt T, Beck A, Cohen MX, Dolan RJ, Wrase J, Heinz A (2010): Prefrontal Cortex Fails to Learn from Reward Prediction Errors in Alcohol Dependence. *J Neurosci*. 30: 7749–7753. [PubMed: 20519550]
28. Stewart JL, Connolly CG, May AC, Tapert SF, Wittmann M, Paulus MP (2014): Striatum and insula dysfunction during reinforcement learning differentiates abstinent and relapsed methamphetamine-dependent individuals. *Addiction*. 109: 460–471. [PubMed: 24329936]
29. Burton AC, Bissonette GB, Vazquez D, Blume EM, Donnelly M, Heatley KC, et al. (2018): Previous cocaine self-administration disrupts reward expectancy encoding in ventral striatum. *Neuropsychopharmacology*. 1.
30. 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. [PubMed: 23752095]
31. 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: 1468–71. [PubMed: 27313048]
32. 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. [PubMed: 25653348]
33. Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hagele C, et al. (2014): Model-Based and Model-Free Decisions in Alcohol Dependence. *Neuropsychobiology*. 70: 122–131. [PubMed: 25359492]
34. 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. [PubMed: 12513946]

Figure 1: Decision-making in the rodent multi-stage 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: grey 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. 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 and the number of days rats spent in each experimental phase presented below.

Figure 2: Model-free behavior predicts addiction-relevant behaviors.

(A) The number of methamphetamine infusions earned in each 6 h self-administration session across the 14 days for individual rats (red lines) and the average of all rats (black dashed line). *** p<0.001 compared to the first day of self-administration. See also Figure S3.

(B) The number of saline infusions earned in each 6 h self-administration across the 14 days for individual rats (gray lines) and the average of all rats (black dashed line). See also Figure S3.

(C) Methamphetamine self-administration data was fit with a power function using maximum likelihood. Drug self-administration data for three individual rats is plotted in the red lines. The number of drug infusions predicted by the fitted power function for these three rats is represented by the black lines. See also 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 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 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. ** $p < 0.01$.
- (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.

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 stage two option ($p(\text{correct} | \text{stage } 1)$) and the probability of choosing the highest reinforced stage two option ($p(\text{correct} | \text{stage } 2)$) in the control/saline 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 stage two option ($p(\text{correct} | \text{stage } 1)$) and the probability of choosing the highest reinforced stage two option ($p(\text{correct} | \text{stage } 2)$) 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. ** $p < 0.01$.

(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. *** $p < 0.001$.

(D) The regression coefficient derived from the logistic regression models in methamphetamine rats before (open bars) and after (closed bars) the selfadministration 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. *** $p < 0.001$.

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 MSDM before (open bars) and after (closed bars) self-administration in control/saline rats. The relationship between each regression coefficient 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 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 relationship between each regression coefficient before and after self-administration for individual rats are presented below the bar graphs with the mean value represented by the red symbol. *** $p < 0.001$. See also Figure S6.

(C) The β_{MB} and β_{MF} estimates obtained from the hybrid RL model characterizing choices in the probabilistic MSDM before (open bars) and after (closed bars) self-administration in control/saline rats. The relationship between each of these parameters before and after self-administration for individual rats is presented to the right of the bar graphs with the mean value represented by the blue symbol. See also Figure S6.

(D) The β_{MB} and β_{MF} estimates obtained from the hybrid RL model characterizing choices in the probabilistic MSDM before (open bars) and after (closed bars) self-administration in methamphetamine rats. The relationship between each of these parameters before and after self-administration for individual rats is presented to the right of the bar graphs with the mean value represented by the red symbol. ** $p < 0.01$; # $p = 0.08$. See also 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 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 Figure S7.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript