Cocaine addiction as a homeostatic reinforcement learning disorder

Mehdi Keramati[1,2,*], Audrey Durand[3,4] , Paul Girardeau[3,4], Boris Gutkin[2,5] , Serge Ahmed[3,4*]

[1] Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, UK.

[2] Group for Neural Theory, INSERM U960, Departément des Etudes Cognitives, Ecole Normale Supérieure, PSL Research University, Paris, 75005, France.

[3] Université de Bordeaux, Institut des Maladies Neurodégénératives, Rue Léo-Saignat, Bordeaux, France.

[4] CNRS, Institut des Maladies Neurodégénératives, Rue Léo-Saignat, Bordeaux, France.

[5] National Research University Higher School of Economics, Center for Cognition and Decision Making, Moscow, Russia.

* Correspondence should be addressed to M.K. (mehdi@gatsby.ucl.ac.uk) or S.H.A. (serge.ahmed@u-bordeaux.fr).

Abstract: Drug addiction implicates both reward learning and homeostatic regulation mechanisms of the brain. This has stimulated two partially successful theoretical perspectives on addiction. Many important aspects of addiction, however, remain to be explained within a single, unified framework that integrates the two mechanisms. Building upon a recently developed homeostatic reinforcement learning theory, we focus on a key transition stage of addiction that is well modeled in animals, escalation of drug use, and propose a computational theory of cocaine addiction where cocaine reinforces behavior due to its rapid homeostatic corrective effect, whereas its chronic use induces slow and long-lasting changes in homeostatic setpoint. Simulations show that our new theory accounts for key behavioral and neurobiological features of addiction, most notably, escalation of cocaine use, drug-primed craving and relapse, individual differences underlying dose-response curves, and dopamine D2-receptor down-regulation in addicts. The theory also generates unique predictions about cocaine self-administration behavior in rats that are confirmed by new experimental results. Viewing addiction as a homeostatic reinforcement learning disorder coherently explains many behavioral and neurobiological aspects of the transition to cocaine addiction, and suggests a new perspective toward understanding addiction.

Keywords: Drug Addiction; Homeostatic Regulation; Reinforcement Learning; Goal-directed Planning; Escalation.

Introduction

Drug addiction is a major public health problem. Decades of experimental research on both human and non-human animal models have revealed the complex, multi-faceted nature of this disorder that several competing theories have attempted to explain, though never comprehensively. One dominant class of theories view addiction as a learning disorder (Everitt & Robbins, 2005; Redish, 2004; Nora D. Volkow & Morales, 2015). According to this view, drugs interfere with the brain's mechanism that updates the rewarding values associated with the experienced actions (Dezfouli et al., 2009; Keramati & Gutkin, 2013; Redish, 2004). Such interference with the value-learning mechanism would, over repeated uses of the drug, results in maladaptive motivation towards drugs due to the over-valued drug-associated choices (through forming rigid drug-seeking habits, or overestimating the value of flexible plans that lead to drugs).

According to another class of influential theories, addiction is viewed as a homeostatic disorder (Serge H Ahmed & Koob, 2005; Newlin, Regalia, Seidman, & Bobashev, 2012). In this view, chronic drug-use induces long-lasting maladaptations in the brain system that is in charge of regulating the organism's critical internal variables to maintain them close to their homeostatic setpoints. Such dysregulation changes the organism's need structure, resulting in escalation of drug use. Escalation of drug taking is a hallmark of the transition from drug use to abuse that precedes the development of more compulsive forms of drug-seeking (S H Ahmed, 2012).

In this paper, we propose an integrative computational theory where escalation of drug use is viewed as a disorder in the brain homeostatic reinforcement learning (HRL) mechanism. The recently proposed HRL theory (Keramati & Gutkin, 2014) takes into account the apparent coupling of the brain reward learning and homeostatic regulation mechanisms (Rangel, 2013). The theory proposes that the rewarding value of an action stems from the anticipated capacity of that action's outcome to reduce the homeostatic deviation of the organism. This computed reward is then used by the associative learning mechanisms as a source of reinforcing associations.

Building upon this framework, our integrative theory incorporates a neurobiologically-plausible minimal model of cocaine's pharmacological effect on several components of the HRL (Keramati & Gutkin, 2014) system. We focus on cocaine as a paradigm example mainly because there is ample evidence, both at the behavioral and neurochemical levels, that its use is regulated by a setpoint-like mechanism (see below). In addition, most preclinical research on escalation of drug use has been conducted with cocaine, generating a rich body of quantitative data to test the validity of our theory. There is currently no comparable data set with other drugs of abuse (Ahmed 2012). However, our theory should be sufficiently general to accommodate other drugs, like methamphetamine, whose consumption is also largely regulated by a set-point-like mechanism.

Heuristically, we argue that the acute administration of cocaine alters the homeostatically-regulated internal state of the organism and thus renders itself a rewarding outcome to be pursued by the organism's goal-directed planning systems. We further propose that chronic use of cocaine induces slow and long-lasting changes in the homeostatic setpoint, equivalent to changing the need structure of the organism. Simulating the model, we replicate and explain potential mechanisms underlying a wide range of behavioral and neurobiological data from rat (S H Ahmed & Koob, 1999; S H Ahmed, 2012; Serge H Ahmed & Cador, 2006; Serge H Ahmed & Koob, 1998; Frantz, O'Dell, & Parsons, 2007; Mihindou, Vouillac, Koob, & Ahmed, 2011; Wee, Specio, & Koob, 2007; Zittel-Lazarini, Cador, & Ahmed, 2007), monkey (Michael A Nader et al., 2006; Michael A Nader, Czoty, Gould, & Riddick, 2008), and human (Nora D Volkow et al., 1999) experiments on cocaine. We demonstrate that several experimental patterns that evade either the purely homeostatic or learning accounts can indeed be explained by a unified HRL approach. We also present new experimental data that confirm a prediction of our theory regarding the exact mechanism by which cocaine intake is regulated, while being incompatible with the predictions of previous homeostatic theories of addiction.

We further frame our theory within the context of previously proposed accounts, pointing out the predictions of those models that are inconsistent with experimental evidence. Particularly, we argue that the habit-based depiction of addiction, for which there exists no direct, unequivocal experimental support, is inconsistent with the specific regularities observed in rats' cocaine self-administration pattern. In contrast, such patterns can be explained by our HRL theory where decisions are made by a goal-directed planning system. Our results also argue that seemingly-habitual addictive behaviors like compulsive drug-seeking are in fact not habitual, but goal-directed behaviors aimed at fulfilling the intensely escalated need for the drug.

Overall, this conclusion is generally consistent with recent research and theoretical thinking about drug addiction in humans that emphasize the importance of goal-directed processes in both the development of and resistance to addiction (Baumeister & Vonasch, 2015; Hogarth & Troisi, 2015; Pickard, 2012) (see Discussion).

Theory Sketch

Homeostatic Reinforcement Learning Framework

Our theory is based on the Homeostatic Reinforcement Learning (HRL) computational framework that is explained extensively elsewhere (Keramati & Gutkin, 2014). Briefly, the HRL theory, like most neuroeconomic theories, postulates that animals learn their environment, and on the basis of their acquired knowledge make choices in order to maximize attained rewarding outcomes. Uniquely to the HRL, when an outcome affects a homeostatically regulated variable, its rewarding value is defined by its ability to fulfill the homeostatic needs of the organism.

Examples of homeostatically regulated internal variables include blood glucose level, body temperature, and plasma osmolality. According to the HRL theory, the rewarding value of an outcome is computed by the extent to which that outcome reduces the distance between the internal variable level and its corresponding setpoint (e.g. by how much eating a cookie would bring the glucose level closer to its setpoint value). In other words, the rewarding value of an outcome is determined by its homeostatic drive-reduction potential. This drive-reduction effect of outcomes can be estimated, possibly erroneously, by their sensory properties like taste and smell.

The HRL framework incorporates this drive-reduction definition of reward into the classical RL theories that provide different neurobiologically-plausible mechanisms for seeking reward (Daw, Niv, & Dayan, 2005; Rangel, Camerer, & Montague, 2008; Sutton & Barto, 1998). Such mechanisms can be organized along two wide classes: habitual and goal-oriented decision processes. The habitual decision process reinforces responses that led to rewards in the past and suppresses responses that led to punishments. In the context of the HRL framework, the habitual process reinforces (or punishes) behaviors that resulted in a reduction (or increase) of homeostatic deviation from setpoint (Keramati & Gutkin, 2014).

Alternatively, a goal-directed decision process exploits an individual's internal model of their environment, i.e., their knowledge of environmental dynamics, to explicitly simulate the consequences that will likely follow a choice. The actual chosen actions come as a result of such explicit deliberation. In the context of HRL framework, the individual assesses a decision-tree expanding into the future to predict the expected outcomes that will ensue a choice, and then estimates the drive-reduction effect of those expected outcomes. In other words, the homeostatic-based primary reward is measured by the reduction in homeostatic deviation anticipated to occur as a result of receiving predicted outcomes. In short, the HRL definition of reward can be normatively incorporated either into the habitual or the goal-directed reinforcement accounts. In this paper, we will argue that several aspects of addiction could arise from a drug-induced disorder of the HRL system that uses goal-directed processes for instrumental decision-making.

Homeostatic Reinforcement Learning Theory of Addiction

Our theory of addiction is based on the critical hypothesis that cocaine acts on the hedonic homeostatic regulation mechanism, and thereby hijacks the reward-computation mechanism. More precisely, we hypothesize that administration of cocaine increases the level of a certain internal variable (hereafter denoted by h_t). Thus, when this variable is below its setpoint level (denoted by h^*), a cocaine infusion results in the reduction of the homeostatic deviation (unless it induces an overshoot) proportional to the self-administered dose. This deviation-reduction renders the cocaine outcome rewarding. We argue later that a plausible candidate for the neurobiological implementation of this variable is the tonic striatal dopamine (DA) concentration, being directly modulated by striatal cocaine concentration (Di Chiara & Imperato, 1988). We will discuss below the details of this implementation (see Neural Substrates).

Fig. 1 lays out the conceptual schema of the theory. The model assumes that upon infusion of the drug, the regulated internal variable h_t first elevates rapidly with increased cocaine level and then falls gradually as cocaine is eliminated from its site-of-action (Fig. 1b). The reinforcement of the cocaine seeking behavior stems form the rewarding value of cocaine, as computed by the anticipated reduction of the homeostatic deviation of h_t from its setpoint (Fig. 1c). Given this anticipated drive-reduction reward, a goal-directed reinforcement learning mechanism (Fig. 1d) control behavior so as to maximize attaining reward.

In other words, a goal-directed planning system, at each point in time, runs a prospective mental simulation of the world and predicts the outcomes (in our case, doses of cocaine) that it expects receive upon choosing every specific course of actions. Given the current internal state, h_t, the model computes the drive-reduction effect of those expected outcomes. This constitutes the rewarding value of that specific course of action. The model then chooses among different options on the basis of their rewarding values.

In parallel with the acute effect of cocaine on the internal state, chronic cocaine use induces several long-lasting neural adaptations, including down-regulation of dopamine D2 receptors availability (Michael A Nader, Daunais, et al., 2002; Michael A Nader, Morgan, et al., 2002; N D Volkow et al., 1993) and reduced dopamine release (N D Volkow et al., 1997) in the striatum. These slow processes can be captured by an adaptive plasticity mechanism where the setpoint, h^*, gradually shifts up in order to compensate for the drug-induced excessive DA concentration (discussed later). In this respect, within our theory, every infusion of cocaine results in a dose-dependent elevation of the setpoint, h^*. Conversely, abstinence gradually lowers the setpoint back to its initial level (Fig. 1c). Thus, as demonstrated in the next section, chronic exposure to sufficiently high doses of cocaine can result in significant elevation of the setpoint. This allows our model to capture several aspects of addiction, including escalation of drug-seeking behavior, elevation of the dose-response curve, and compulsive drug-seeking.

In order to capture neurobiological constraints on setpoint level (i.e., biological variables cannot take infinite values), we set a lower-bound ($\underline{h}^*$) and an upper-bound ($\overline{h}^*$) on the setpoint level (Fig. 1c). That is, cocaine exposure can gradually elevate h^* to a maximum level $\overline{h}^*$, and long-term abstinence can gradually lower the setpoint level to a minimum level $\underline{h}^*$. In all simulations in this paper we assume that before exposure to cocaine, animals have an initial setpoint level, h^*, equal to the lower-bound, $\underline{h}^*$. The only exception to this is when we analyze individual differences and assume different initial setpoint levels for different animals.

In sum, our computational theory assumes that cocaine reinforces behavior due to its rapid effect on the internal state (i.e., drive-reduction effect), whereas its chronic use induces slow and long-lasting changes in homeostatic setpoint.

Simulation of Empirical Results

Escalation and Changes in Loading Patterns

There is now strong evidence showing that following a history of extended access to cocaine self-administration (SA), rats present behavioral changes that recapitulate important behavioral features of cocaine addiction. A seminal experiment (S H Ahmed & Koob, 1999; Serge H Ahmed & Koob, 1998) demonstrated that with one hour access per session (short access or ShA) to intravenous cocaine SA, drug intake remained low and stable. In contrast, with 6 hours of access (long access or LgA), drug intake gradually escalated over days and eventually reached a level 200% greater than that of ShA rats.

Consistent with the replicated experiments (S H Ahmed & Koob, 1999; S H Ahmed, 2012; Serge H Ahmed & Cador, 2006; Serge H Ahmed & Koob, 1998, 2005; Hao, Martin-Fardon, & Weiss, 2010; Mihindou et al., 2011; Piazza, Deroche-Gamonent, Rouge-Pont, & Le Moal, 2000; Wee et al., 2007; Zittel-Lazarini et al., 2007), we simulated our model in a virtual task (Fig. S1) where each lever-press (fixed-ratio one) initiated an intravenous infusion of cocaine over 4sec, followed by a 20sec time-out period during which the lever was inactive (**Fig. 1**a). It is important to note that all experimental data presented in this paper was replicated by simulating the model with one single set of values for the free parameters (Table. S1).

Model simulation results (Fig. 2) replicate the experimental data (S H Ahmed & Koob, 1999; Serge H Ahmed & Koob, 1998) (Fig. S2, Fig. S3). At the start of each simulated session the simulated agents start in a cocaine-deprived state (internal state = h_t = 0), hence the internal state is far below the setpoint. Thus, the cocaine injection outcome can reduce homeostatic deviation and therefore is rewarding. This results in an initial burst of responding (known as "loading") that allows the setpoint to be reached at the beginning of each session. After reaching the setpoint, the agents settle to a steady level of responding that maintains homeostasis (note that as in the replicated experiments, the structure of the task was learned by the model during a pre-training period).

Furthermore, our simulations fully replicate the pattern of the infusion-rate escalation in the LgA animals and make apparent that this escalation could be due to the gradual elevation of the setpoint over several 6-hr access daily sessions. In ShA animals, however, the setpoint elevation during every 1-hr session is small enough that the rest of the day (23-hrs) is sufficient for a full recovery to the initial setpoint level. In the LgA, the setpoint does not recover to the initial level. Such persistently elevated homeostatic setpoint in the LgA agents induces the escalation because maintaining the internal state at a higher setpoint requires more infusions to compensate for the relatively faster elimination of cocaine at the higher concentrations. Clearly, at higher concentrations, a constant elimination rate (or half-life) naturally leads to higher total amount of eliminated cocaine.

Pre- and Post-escalation Dose-response Curves

Experiments have shown that post-escalation infusion rate, as well as the total amount of consumed cocaine per hour, are higher in the LgA than in the ShA animals, for all unit doses of cocaine. However, in both the LgA and the ShA animals, whereas the infusion rate decreases as a function of dose, the amount consumed varies little (S H Ahmed & Koob, 1999; Serge H Ahmed & Koob, 1998) (Fig. S4).

Our simulation results replicate these patterns (Fig. 3a, b). The decreasing dose-response curve in our simulations is due to the fact that at the higher doses, a single infusion can keep the internal state at a high level (i.e., close to the setpoint) for a relatively longer time, as compared to the lower doses. Therefore, at high doses, a smaller number of infusions is required to maintain the homeostasis (Fig. 3a). Although the number of infusions decreases as a function of dose, the total cocaine intake remains stable (Fig. 3b), since total intake is the product of the infusion rate and the unit-dose.

Post-escalation Reduced Availability

Evidence shows that post-escalation reduction of access duration from 6hr to 1hr per day in the LgA rats results in a gradual decline of the infusion rate. This decline becomes faster when the access to cocaine is limited even further to only 1hr per week (S H Ahmed & Koob, 1999) (Fig. S5). According to our simulation results, shortened access allows the escalated setpoint to gradually return to its initial level (Fig. 3d), resulting in a gradually decreasing infusion rate (Fig. 3c). The 1hr weekly access speeds up this process even further (Fig. 3c, d), since it decreases the exposure to the drug and weakens the setpoint elevation effect of cocaine. In parallel, the increased duration of periods without cocaine enhances the recovery process of the setpoint (i.e., the return of the setpoint back to its initial level). Notably, results further predict that five weeks of limited access is still insufficient for complete recovery to the initial, pre-escalation setpoint (Fig. 3d) (further discussed in the section “testable predictions”).

Effect of Session-duration on Escalation

Given that access-duration is a continuous experimental parameter, one might expect that decreasing this parameter will result in smoothly weaker escalation in rats. Yet strikingly, experiments revealed a minimum duration of access to the drug, below which no escalation is possible and above which the speed of escalation increases with the duration of access (Wee et al., 2007) (Fig. S7). Our model naturally accounts for the existence of such critical duration. In fact, increasing the session duration prolongs the daily elevation and shortens the daily recovery periods of the setpoint level and thus, accelerates escalation (Fig. 4). For very short session durations (e.g. 1hr), however, the recovery effect during the rest of the day (e.g.

23hrs) completely cancels out the small elevation during the SA session. Above a certain critical duration, the setpoint up-regulation is no longer canceled out, hence escalation emerges. For session durations that are above this critical duration, increasing the duration monotonically speeds up escalation.

The model also predicts that at this critical access-duration (3hr, in our simulation results), the recovery period (that is 21hr, in our simulation) is just enough to cancel the cocaine-induced setpoint elevation (Fig. 4d). Hence, as a testable prediction, the model argues that such an access-duration will not induce escalation (Fig. 4), but will preserve the infusion rate at an escalated level (e.g. if it has been escalated previously, using a 6hr access condition). Further details of this prediction are explained later in the section “Testable Predictions”.

Punishment-induced Suppression of Drug-Seeking

Experimental results also show that post-escalation pairing of cocaine infusion with electric shock results in rapid suppression of responding within the first 45-minute session in both ShA and LgA rats (S H Ahmed, 2012). When the punishment is removed after this session, whereas the LgA animals rapidly resume infusion with the pre-punishment rate within the first 45-min session, the infusion rate in the ShA rats remains suppressed even after three sessions (S H Ahmed, 2012) (Fig. S9). Such resistance to the long-term effects of punishment in the LgA rats is supposed to capture compulsive drug seeking, which is the hallmark of addiction in humans.

Simulating the model reproduces these data (Fig. 5), where the effect of electric shock is modeled by attributed a fixed, internal state-independent, punishment (i.e., negative reward) to each administration of the electric shock (Fig. S8). According to the model, the high cost of lever-press during the electric-shock phase suppresses the agents’ motivation to defend homeostasis. That is, the agents only seek cocaine when the internal state falls far below the setpoint and thus, the deviation-reduction reward of cocaine outweighs the fixed punishment of electric shock. Behaviorally we observe that the agents significantly reduce their response rate (Fig. 5a). During the post-punishment sessions, although the agents still expect punishment upon pressing the lever, the homeostatic deviation in the agents whose setpoints have been persistently elevated (LgA) is just large enough to motivate occasional lever-presses. Through this occasional exploratory behavior, the LgA agents get the opportunity to learn that the punishment is removed and therefore, resume SA with the pre-punishment rate. The ShA agents have a relatively lower “need” for cocaine and hence, their chances to explore the punishment-removed environment are much reduced. Therefore, as with experimental results (S H Ahmed, 2012), a fraction of such agents never try even a single lever-press, whereas some others try and thus, learn that the punishment is removed. As a result, and consistent with experimental data (S H Ahmed, 2012) (Fig. S7), the ShA agents on average, but with a high inter-individual variability, continue cocaine-seeking at a suppressed rate (Fig. 5b).

Extinction and Priming-induced Relapse

Another important aspect of drug-addiction is the pattern of drug craving and relapse that occurs even after long-term abstinence. Well-validated animal models of craving and relapse show drug-primed reinstatement of cocaine seeking following extinction and abstinence. For example Mihindou et al. (Mihindou et al., 2011) performed an experiment where, following 32 sessions of 6hr access to cocaine, rats underwent 10 days of reinstatement procedure. Each day consisted of 5 consecutive 45min blocks during which pressing the lever had no consequence (extinction). At the beginning of each block, the rats received a single priming injection of cocaine with the following doses: 0, 0, 0.25, 0.5, and 1mg.

Simulating the model (Fig. 6) in these same conditions reproduced the patterns observed in experimental results (Mihindou et al., 2011) (Fig. S9). Firstly, as in behavioral data, although drug-seeking was extinguished in the first two blocks (no cocaine) of the first day, rats relapsed dose-dependently and transiently after each priming injection, and was followed by a gradual return to pre-priming levels of cocaine seeking (Fig. 6a, curve Pr1; also see Fig. S9c, curve Pr1 for similar experimental results). We can explain the underlying processes that rise to this pattern in our model as follows.

In the HRL framework, the world in which the organism lives is composed of both external and internal states. Changes in the drug-controlled internal state, h_t, aside from its rewarding effects (due to reducing homeostatic deviation), also act on the agent's state-recognition system (Gershman, Blei, & Niv, 2010; Redish, Jensen, Johnson, & Kurth-Nelson, 2007), analogous to the role of contextual states (Fig. S1). The state-recognition system allows the organism to recognize that it is in the "under-cocaine" or "cocaine-free" situation depending on its internal state. This mechanism is supported by many studies showing that cocaine, like other drugs of abuse, acts as an interoceptive and discriminative cues (Desai, Paronis, Martin, Desai, & Bergman, 2010; Lamas, Negus, Gatch, & Mello, 1998; Mantsch & Goeders, 1998). As a result, having been extinguished in a drug-free state of the world (i.e., when h_t is low) during the first two extinction blocks, the association between lever-pressing and cocaine has remained intact in the "under-cocaine" state (i.e., when h_t is high) (Fig. 6e, session 1). In other words, the agent still expects cocaine in the "under-cocaine" state. Therefore, since priming injection re-induces the interoceptive stimuli of cocaine by temporarily increasing h_t (Fig. 6f), drug-seeking relapses (Fig. 6b, curve Pr1). As the priming cocaine starts degrading, the agent gradually returns to the "drug-free" state (Fig. 6f) where drug-seeking had been previously extinguished. This explains the transient nature of the priming effect on cocaine seeking (Fig. 6c, curve Pr1; also see Fig. S9c, curve Pr1 for similar experimental results).

Moreover, consistent with experimental results (Mihindou et al., 2011) (Fig. S9c), when the priming dose is sufficiently high (1mg), the reinstatement does not occur instantaneously after the priming injection, but peaks with a delay of 5-10min (Fig. 6c, doses 0.5 and 1 mg). According to our simulations, this is because upon injection of a high dose, the internal state overshoots the setpoint (Fig. 6f). Although this overshoot still results in the agent entering the “under cocaine” state where the association between lever-press and outcome is strong, the drug outcome in this situation is not rewarding as it will only increase the homeostatic deviation even further. Thus, the agent waits until the internal state drops sufficiently below the setpoint and then starts seeking cocaine (Fig. 6c).

Priming-induced reinstatement is in fact a transient phenomenon, and gradually extinguishes over approximately 10 days of experiencing the reinstatement procedure (Mihindou et al., 2011) (Fig. S9a, b). According to our possibly counterintuitive simulations results, the extinction of the cocaine-induced reinstatement is not due to a gradual recovery of the setpoint to its pre-addiction level (even after 10 days, Fig. 6d), but due to gradual extinction of lever-cocaine associations in the “under-cocaine” state (Fig. 6e). This happens after the agent sufficiently experiences lever-presses with no cocaine, under the effect of priming cocaine.

Effect of Session-duration on Priming-induced Relapse

Experimental and simulation results in the previous subsection only focused on priming-induced relapse in LgA animals. However, comparing the reinstatement patterns between the ShA and LgA animals, having a history of long access to the drug leads to a more pronounced cocaine-induced reinstatement of the drug-associated behaviors across all tested doses (Serge H Ahmed & Cador, 2006) (Fig. 7; also see Fig. S10 for similar experimental results). According to our model, this is because the elevated setpoint level in the LgA agent induces a stronger homeostatic deprivation (deviation) and thus, results in a higher estimated value for the cocaine outcome. This, in turn, results in more pronounced reinstatement in the LgA agent (Fig. 7).

Dose-response Curve and Individual Differences

According to our model, when the unit-dose of drug is very low, even the small cost of drug-seeking (e.g., pressing lever) could outweigh the drive-reduction rewarding effect of the drug. In such cases (First three rows in Fig. 8), the rate of lever-press stays at zero (Fig. 8, left column) and consequently, the internal state stays at the drug-free level (Fig. 8, right column).

By further increasing the unit-dose, the agents start self-administering the drug when the internal state is very low. When the internal state rises above the drug-free level, however, the agents’ need for drug is partially fulfilled. At this point the rewarding drive-reduction effect associated with seeking further drug decreases (as compared to the drug-free case), possibly to a level the small cost associated with obtaining the drug outweigh it. As a result, for these intermediate unit-doses (4th and 5th rows in Fig. 8), the internal state deviates from the drug-free level, yet the lever-pressing rate is not sufficiently high to reach the setpoint. Based on these observations, we can define the “critical dose” as the smallest unit-dose at which the agents start pressing the lever for the drug (e.g., the 4th row in Fig. 8).

Increasing the unit-dose above the critical dose, we get to a point where the unit-dose becomes sufficiently high so that the internal state can reach the setpoint (6th row in Fig. 8). This level of unit-dose produces the "maximum infusion rate". This is because at higher unit-doses, the agents must decrease their rate of infusion in order to avoid systematically overshooting the setpoint. In other words, at sufficiently high doses (e.g., 9th and 10th rows in Fig. 8), the infusion rate must be sufficiently low in order to provide enough time for the high brain concentration of the administered cocaine to decrease.

In sum, our model predicts that increasing the unit-dose produces a dose-response curve that has zero infusion rate for very low dose, followed by a sharp rise, and then followed by a gradual decrease (Bottom left panel in Fig. 8). This is equivalent to the pattern reported in animal models of cocaine self-administration (Zittel-Lazarini et al., 2007) (Fig. S11). However, experimental results also show that different animals have different "critical doses", as well as different "maximum infusion rates" (Zittel-Lazarini et al., 2007) (Fig. S11). These individual differences can be explained by simulating our model under the assumption that different individual agents have different setpoint levels (in our simulations, we simply attribute different setpoint levels to different agents). These individual setpoint levels results in different "critical doses", as well as different "maximum infusion rates" (Fig. 9a-f). This mechanism makes the strong prediction that there is an inverse correlation among animals between the "critical dose" and "maximum infusion rate". The next section explains and tests this prediction.

Tested predictions

Inverse Correlation between Critical Dose and Maximum Infusion Rate

Our theory predicts that the critical unit-dose at which animals start seeking cocaine will be lower in animals with higher setpoint levels (vertical dashed lines in Fig. 9a-f). This is because even a small unit-dose of cocaine has a strong deviation-reduction effect for the animals with a high setpoint. In fact, having a higher setpoint is equivalent to having a higher deprivation level. Deprivation level, in the HRL theory, has an excitatory effect on the deviation-reduction reward (Keramati & Gutkin, 2014), meaning that the larger is the deprivation level, the more "rewarding" is the same unit-dose of the drug.

Our model also predicts that the higher the setpoint level is, the higher the maximum infusion rate among the different unit-doses will be (horizontal dashed lines in Fig. 9a-f). This is again due to faster elimination of cocaine at the higher levels (due to higher setpoints that are reached), which requires higher infusion rates to compensate for it.

In sum, our theory proposes that the individual differences in dose-response curves (Zittel-Lazarini et al., 2007) stem from differences in the setpoint levels. The theory thus predicts that by increasing the setpoint level, the critical dose decreases, but the maximum infusion rate increases. This predicts a behaviorally-observable inverse correlation between the critical dose and the maximum infusion rate (Fig. 9g). To test this prediction, we analyzed previously published data from rats (n=17) self-administering different doses of cocaine (Zittel-Lazarini et al., 2007), and revealed that this inverse correlation is in fact significant ($p < 0.01$) in experimental evidence (Fig. 9h).

Effect of Within-session Reduction of Unit-dose

In our theory decision making is geared toward maximizing the collected rewards. This constrains the objective of the agent to maintain the internal state as close as possible to the setpoint. To this end, the agent self-administers cocaine with a stable rate so that the internal state fluctuates regularly "around" the setpoint (Fig. 10c, d). Thus, a self-administration response is triggered each time the internal state drops "sufficiently" below the setpoint; that is, when the setpoint is between the current internal state and the state that will be reached after taking one dose (Fig. 10c, d). This is in contrast to the previous regulatory models of cocaine addiction (Serge H Ahmed & Koob, 2005; Newlin et al., 2012; Tsibulsky & Norman, 1999) where a response is triggered as soon as the internal state just drops below the setpoint (Fig. 10a, b). Therefore, those models predict that for different unit doses of cocaine, the response-triggering state should be the same (equal to the setpoint) (Fig. 10a, b), while our model predicts different triggering states for different unit doses (Fig. 10c, d). That is, the smaller the dose is, the closer to the setpoint the internal state will be maintained and thus, the higher the minimum level of cocaine concentration (or striatal DA level) will need to be.

Testing these diverging predictions in a direct way would require measurements of cocaine or DA levels immediately before each triggered response (Fig. 10). However, they can nevertheless be tested using a rather simple behavioral experiment involving a large, non-signaled within-session decrease in the unit dose of cocaine (e.g., from 1 to 0.0625 mg per injection). In this case, our model predicts a stable low-rate response when the unit-dose is high, followed by a loading burst just after reducing the dose, and eventually a stable high-rate response (Fig. 11c). Previous regulatory models of cocaine addiction (Serge H Ahmed & Koob, 2005; Newlin et al., 2012; Tsibulsky & Norman, 1999), in contrast, predict no loading phase after dose reduction, but a simple transition from a stable low-rate response to a stable higher-rate response (Fig. 11g).

In our model, the reason for the loading burst after the dose reduction is that the agent, when the dose has just decreased, is initially unaware of this change. Thus, as if nothing has changed, the agent waits until the internal state drops sufficiently below the setpoint. Upon the first post-dose-reduction infusion, the agent learns (either fully, or partially due to a learning rate) that the dose is reduced and that taking it is not enough to reach the setpoint. As a result, the agent initiates responding at the highest possible rate until the setpoint is reached (loading phase). This burst of responding is then followed by a stable and relatively higher response rate that guarantees that the internal state fluctuates around the setpoint with the new reduced dose (Fig. 11a-d). Previous models, in contrast, predict no loading phase after dose reduction, since the internal state is always maintained above the setpoint and thus, there is no undershoot to be compensated for when the dose is reduced (Fig. 11e-h). In summary, our theory predicts that the first few responses after dose-reduction would happen at a higher rate (i.e., the loading phase), as compared to the stable response-rate that will follow. In other words, it predicts that the inter-infusion intervals (III) for the first few responses after dose-reduction would be smaller compared to the subsequent IIIs.

We tested this behavioral prediction in rats (n=21) trained in 2-hour sessions of cocaine SA with unit doses of 1 and 0.125 mg/injection during the first and the second hours, respectively. We assumed that animals' response rate reaches its steady state after 10 trials following dose-reduction. Thus, for each animal, we computed the average III between the 10th and the 20th post-reduction responses (20 was the maximum number of post-dose-reduction responses among all the animals). This supposedly stable rate was then compared to the IIIs just after dose-reduction. Our experimental results showed that in fact, the first two post-dose-reduction IIIs were significantly shorter than IIIs between the 10th and the 20th post-reduction responses (Fig. 11i, j) (See Supplementary Materials for details of the experiment and statistical analysis). This dose reduction-induced loading phase verifies our prediction and contradicts the prediction of previous models.

Testable Predictions

Different Satiety Thresholds for Different Unit-doses

As explained in the previous section, our theory predicts that the response-triggering internal state would be different for different unit doses of cocaine (Fig. 10c, d). This prediction depends on the critical assumption of the model that in order to minimize the overall homeostatic deviation over time agents strive to keep the internal state "around" the setpoint,. This is in contrast to the previous regulatory models of cocaine addiction (Serge H Ahmed & Koob, 2005; Newlin et al., 2012; Tsibulsky & Norman, 1999) that predict equal response-triggering internal states for different unit doses (Fig. 10a, b).

Time-out Effect on Loading and Pause Patterns

As discussed before, our model explains the experimental evidence (Serge H Ahmed & Koob, 2005) that cocaine self-administration sessions start with a loading period during which animals show a burst of cocaine-seeking responses, and then followed by a pause period during which no injection is taken. According to our model, the loading pattern is a behavioral strategy for changing the internal state from the initial drug-free state to the homeostatic setpoint, as fast as possible. Due to the short delay between each infusion and its full effect on the internal state (the ascending limb of the curve in Fig. 1a), the agents keep taking cocaine for some excessive times during the loading phase, even after having taken enough for reaching the setpoint. This results in overshooting the setpoint and thus, a pause period so that this excess cocaine degrades or washes out.

In this respect, the model predicts that both loading and pause patterns will be more pronounced by decreasing the timeout period (the period after each infusion when the level becomes inactive) (Fig. 12). Intuitively, a shorter timeout (e.g., 4 seconds) provides the agent with the opportunity of adopting a higher infusion rate during the loading period. This increases the number of excessive infusions during the loading phase, which in turn, results in a stronger overshoot and thus, a longer pause for offsetting it. A longer timeout (e.g., 20 seconds), in contrast, provides more time for each administered dose to affect the internal state. Therefore, the agent has a more realistic measure of its internal state before taking the next dose. This reduces the chance of overshooting the setpoint and thus, predicts a shorter pause period following the loading phase.

It is noteworthy that loading and pause patterns, as well as the prediction offered in this subsection, stem from the homeostatic/pharmacological components of our model (i.e., dynamics of drug-absorption and washout, setpoint level, etc). Thus, for the case of these patterns, our model has equal explanatory power as compared to previous pharmacological models of addiction (Serge H Ahmed & Koob, 1998, 2005; Pickens & Thompson, 1968).

Escalation-preservation under a Critical Session-duration

As explained above, we predict that there would exist a critical daily access-duration (3hr, in our simulation results) where the daily recovery period (i.e., 21hr, in our simulation) is just enough to cancel the cocaine-induced setpoint elevation (Fig. 4d). Under such access-duration condition, the setpoint neither escalates, nor recovers back to its pre-exposure level. Instead, it will have a steady level over time.

Hence, as a testable prediction, the model argues that such an access-duration would preserve the infusion rate at its current escalated level (Fig. 13). That is, if an escalated animal (e.g. escalated under a 6hr condition) experiences the 3hr condition for a some days or weeks, its setpoint and thus its response rate would stay at a stable level during this time. On the other hand, no escalation would take place if a short-access animal is provided this 3 hour access.

Rapid Re-escalation after Priming-based Extinction

As explained before, according to our simulations results (Fig. 6), the extinction of cocaine-induced reinstatement is not due to a recovery of the setpoint to its pre-addiction level (even after 10 days, Fig. 6d), but due to a gradual extinction of lever-cocaine associations in the “under-cocaine” state (Fig. 6e). This happens after the agent experience sufficient lever-presses without cocaine outcome, under the effect of priming cocaine.

Thus, of critical clinical importance, we predict that once rats are again given access to cocaine SA after the 10-day extinction procedure, the infusion rate must return rapidly (within only one 6hr session) to its escalated level. This is because the setpoint remains at an elevated level, and the extinguished lever-cocaine associations can be re-learned within less than one hour (Fig. 14c), resulting in rapid re-escalation of infusion rate (Fig. 14a, b).

Neural Substrates

Our model raises the important question of the neural parameters that are regulated during cocaine self-administration. We know that during cocaine self-administration, animals rapidly achieve and then maintain dopamine within the nucleus accumbens (NAc) at a high, albeit relatively steady, level (Pettit & Justice, 1989; R A Wise et al., 1995). This self-maintained level of dopamine is considerably increased after escalation of cocaine self-administration (Serge H Ahmed, Lin, Koob, & Parsons, 2003). A more fine-grained temporal analysis of NAc dopamine fluctuations has also revealed that rats are more likely to self-administer an additional dose of cocaine when dopamine drops by a certain amount below the high level maintained during cocaine self-administration. Inversely, they are unlikely to self-administer an additional dose when dopamine is above that level (R A Wise et al., 1995). These in vivo neurochemical findings suggest that tonic NAc dopamine level could be one of the major neurobiological variables regulated during cocaine self-administration.

However, it is unlikely that tonic NAc dopamine level per se is directly regulated, but instead its postsynaptic effects on the activity and/or excitability of NAc neurons (see below). This is corroborated by seminal findings showing that pharmacological treatments that selectively increase or decrease NAc dopamine receptor signaling reduce or increase the rate of cocaine self-administration, respectively (Bachtell, Whisler, Karanian, & Self, 2005; Bari & Pierce, 2005; Caine, Heinrichs, Coffin, & Koob, 1995; Suto, Ecke, & Wise, 2009; Suto & Wise, 2011; Thanos, Michaelides, Umegaki, & Volkow, 2008). In addition, this is supported directly by in vivo multiple single-neuron recordings in the NAc of cocaine self-administering rats (L L Peoples, Gee, Bibi, & West, 1998; Laura L Peoples & Cavanaugh, 2003; Laura L Peoples, Kravitz, Lynch, & Cavanaugh, 2007; Laura L. Peoples, Uzwiak, Gee, & West, 1999) (reviewed in (Laura L. Peoples, Kravitz, & Karine, 2010)). These elegant studies have consistently shown that the firing activity of most NAc neurons is either tonically decreased (about 60%) or increased (about 40%) during cocaine self-administration. NAc neurons show also phasic changes in firing activity time-locked to instrumental responding for cocaine (e.g., lever pressing) but these fluctuations are smaller than the average steady change maintained throughout drug self-administration (Laura L. Peoples et al., 2010; Wheeler & Carelli, 2009). What is the overall effect of these opposite self-maintained changes in tonic activity between different NAc neurons?

Most NAc medium spiny neurons (MSN) are GABAergic output neurons that project downstream to either the same (converging) or different (diverging) brain regions, such as the ventral pallidum and the lateral hypothalamus (Scofield et al., 2016; R. J. Smith, Lobo, Spencer, & Kalivas, 2013; Zahm, 2000). Thus, one plausible regulated parameter, though not necessarily the only one, during cocaine self-administration is the balance of neuronal output activity within or between distinct NAc downstream pathways. For instance, different NAc neurons send projections that converge to the ventral pallidum – a brain region critically involved in hedonic processes related to both drug and nondrug rewards (Root, Melendez, Zaborszky, & Napier, 2015; K. S. Smith, Tindell, Aldridge, & Berridge, 2009). There, their activities can be integrated into a single variable that can serve as input to a setpoint-based regulation mechanism. This general hypothesis is similar to a previous proposal in (Carlezon & Thomas, 2009) in that it ascribes to NAc neuronal activity a critical role in reward in general and in drug reward in particular. However, our hypothesis differs in that it attributes equivalent importance to NAc neuronal excitation and inhibition in the regulation of drug reward during self-administration.

What distinguishes NAc neurons whose firing activity is tonically increased or decreased during cocaine self-administration? At this stage, we can only speculate from incomplete data, obtained mainly outside a drug self-administration setting. Dopamine acts on both D1-like and D2-like receptors that are expressed by different, largely segregated, populations of neurons in the nucleus accumbens, like in other striatal territories: the D1R and D2R neurons (Gerfen & Surmeier, 2011; Scofield et al., 2016). Dopamine has opposite effects on the excitability and/or activity of NAc D1R and D2R neurons: it increases activity and/or excitability of D1R neurons but decreases activity and/or excitability of D2R neurons (Gerfen & Surmeier, 2011). Consistent with this, many studies using a variety of methods have consistently shown that cocaine increases activity of D1R neurons but reduces activity in D2R neurons, thereby shifting NAc neuronal output in favor of D1R neuronal activity (Bertran-Gonzalez, Laurent, Chieng, Christie, & Balleine, 2013; Calipari et al., 2016; Chandra et al., 2013; Luo, Volkow, Heintz, Pan, & Du, 2011; Park, Volkow, Pan, & Du, 2013). The increase in activity of D1R neurons may result not only from a direct effect of cocaine-induced dopamine on their excitability (Lüscher & Malenka, 2011) but also from a major indirect disinhibition effect that involves an inhibition of D2R neurons' collaterals that inhibit D1R neurons (Bock et al., 2013; Dobbs et al., 2016). On the basis of this information, we speculate that NAc neurons whose activity is tonically decreased during cocaine self-administration would mainly correspond to D2R neurons while NAc neurons whose activity is tonically increased would mainly correspond to D1R neurons. Thus, the goal that animals would pursue during cocaine self-administration is to maintain a sustained increase in NAc D1R neuronal output activity via a sustained high dopamine level. This hypothesis is consistent with the well-known role of NAc D1R neurons in reward in general (Lobo & Nestler, 2011; Soares-Cunha, Coimbra, Sousa, & Rodrigues, 2016; Yager, Garcia, Wunsch, & Ferguson, 2015) and in drug reward in particular (Hikida, Kimura, Wada, Funabiki, & Nakanishi, 2010; Lobo et al., 2010). Whether this increased output activity affects all downstream NAc D1R neuronal pathways in parallel or only one specific pathway (e.g., the Shell NAc-VP pathway) is difficult to establish at present.

Within this admittedly preliminary model, escalation of cocaine self-administration may result from a chronic decrease in NAc D1R neuronal activity. Animals would increase cocaine intake to maintain NAc dopamine at a higher level than prior to escalation (Serge H Ahmed et al., 2003) in an attempt to compensate for a decrease in NAc D1R neuronal activity (Serge H Ahmed & Koob, 2004; Doyle et al., 2014; Self, 2014). Basal decrease in NAc neuronal activity has been observed after both escalation and post-abstinence re-escalation of cocaine self-administration, particularly in the NAc shell subdivision (Guillem, Ahmed, &

Peoples, 2014). However, whether this decrease in activity preferentially affects NAc DR1 neurons, though likely, has yet to be demonstrated directly. Nevertheless, moving forward, we can briefly envision several neuroadaptive mechanisms that could contribute to explain a decrease in basal D1R neuronal activity. First, it may result from a low basal dopamine level that consequently stimulates less than normal the excitability and/or activity of NAc D1R neurons. However, no decrease in basal dopamine has so far been observed after escalation of cocaine self-administration (Serge H Ahmed et al., 2003). Second, a decrease in NAc D1R activity may also result from a decrease in excitatory glutamatergic inputs onto D1R neurons or a change in membrane intrinsic excitability (Kourrich, Calu, & Bonci, 2015). There is some evidence that reduced NAc glutamate levels after chronic cocaine self-administration can contribute to some aspects of cocaine-seeking behavior (Baker et al., 2003; O. M. Ben-Shahar et al., 2012; Kalivas, 2009; Scofield et al., 2016) but its role in escalated levels of cocaine self-administration remains to be fully demonstrated (Doyle et al., 2014). Interestingly, after days or weeks of abstinence from cocaine self-administration, there is a potentiation of glutamatergic inputs onto NAc D1R neurons (Pascoli et al., 2014; Terrier, Lüscher, & Pascoli, 2016; Wolf, 2016) that coincides with a de-escalation of cocaine intake and an incubation of cocaine craving (Guillem et al., 2014). Whether this potentiation of glutamatergic inputs causes de-escalation via an increase in basal NAc D1R neuronal activity, as hypothesized here, remains to be demonstrated.

Finally, a decrease in NAc D1R neuronal activity could also arise from an increased local inhibition, notably from the subset of NAc D2 neurons that provide lateral inhibition of neighboring D1R neurons. The latter could be due in turn to a decrease in D2 receptor function and/or number in NAc D2R GABAergic collaterals, thereby making these inhibitory collaterals less susceptible to dopamine inhibition (Dobbs et al., 2016). Interestingly, the latter mechanism, though highly speculative at present, is nevertheless consistent with reduced striatal D2 receptor availability in people with cocaine addiction (Martinez et al., 2009; N D Volkow et al., 1993) and in nonhuman primates after extended cocaine self-administration (Michael A Nader et al., 2006, 2008). However, such decrease has not been consistently found in the NAc of rats after escalation of cocaine self-administration (O. Ben-Shahar et al., 2007; Briand et al., 2008; Conrad, Ford, Marinelli, & Wolf, 2010).

According to our model, the down-regulation of D2R, that in our framework would be equivalent to elevation of the setpoint, should result in a higher homeostatic deviation under normal (i.e., non-drug) conditions. This, in turn, would lead to a higher rewarding value (i.e., drive-reduction effect) associated with drug consumption (Fig. 15). If we presume that the drive-reduction impact of the drug also has a positive hedonic valence, this scheme may explain the inverse correlation between D2R availability and the reported pleasantness of taking psychostimulants, observed in human addicts (Nora D Volkow et al., 1999) (Fig. S12).

D2R availability is also inversely correlated with motivation for drugs under drug conditions, as measured by steady-state rate of cocaine SA in monkeys (Michael A Nader et al., 2006)

(Fig. S13). According to our model, the lower D2R availability (i.e., escalated setpoint) motivates the animal to maintain the striatal cocaine concentration at an elevated level (Serge H Ahmed et al., 2003). Due to the faster elimination of cocaine at higher levels, however, defending homeostasis requires a higher infusion rate (Fig. 16), explaining the increased motivation for drugs in monkeys with lower D2R levels.

Though much remains to be done to incorporate other important behavioral and neurobiological details into our setpoint model, it nevertheless defines tonic NAc D1R neuronal output activity as a plausible regulated variable during cocaine self-administration. This hypothesis is generally consistent with previous research and theoretical interpretation that has attributed a predominant, though not necessarily exclusive, role of NAc D1 receptors and D1R neurons in the satiating or primary rewarding effects of cocaine (Lobo & Nestler, 2011; Self, 2014). In addition, it delineates several neuroadaptive mechanisms by which the value of the regulated variable can be altered by extended drug access to induce escalation of cocaine self-administration (e.g., decreased D2 receptor availability in D2R neurons). Finally, this model also predicts that manipulating the activity of NAc D1R neurons, either directly or indirectly via NAc D2R neurons, should affect escalation of cocaine self-administration.

Discussion

In this paper, we showed that several aspects of cocaine self-administration in rats, notably escalation with extended access, could be parsimoniously explained as drug-induced disorders in a homeostatic reinforcement learning system. Those explained patterns arise from several mechanisms embedded in the model, or from their interaction. The basic rewarding effect of cocaine results from its effect on moving the internal state closer to a homeostatic setpoint, as long as it has not overshot the setpoint. This mechanism gives rise to the loading effects at the beginning of self-administration sessions or after a within-session dose-reduction, as well as the subsequent pause effect, followed by the steady rate of infusion for keeping the internal state around the setpoint. Furthermore, it produces a decreasing dose-response curve, but a dose-independent total-consumption curve.

Further critical mechanism in the model is the gradual long-lasting effect of cocaine on elevating the setpoint level. This mechanism results in higher steady-state infusion rates as the setpoint elevates (and consequently an upshift of the dose-response curve), because keeping the internal state at an elevated level requires more infusion as compared to when the setpoint is not elevated. This is simply due to the pharmacological principle that circulating cocaine (or any other molecule) degrades/washes-out faster when it is at higher levels. Furthermore, the speed of drug-induced setpoint elevation and its recovery speed brings about the effect of session-duration on escalation pattern, as well as gradual extinction during abstinence.

The interaction of the reward-generating mechanism of cocaine and the setpoint elevation mechanism, leads to compulsive drug-seeking in LgA, but not ShA agents as measured by resistance of cocaine-seeking and taking to foot shocks. Furthermore, their interaction with the instrumental conditioning mechanism that learns the association between lever-press response and cocaine outcome leads to the initial acquisition (learning) of the drug-seeking behavior and results in the priming-induced relapse pattern and its extinction.. The necessity of using a goal-directed reinforcement learning algorithm for this instrumental conditioning mechanism is discussed in the following subsection.

The validity of our computational model was tested on raw self-administration data obtained from experiments conducted by Serge Ahmed and his co-workers. Hence, while in this work we focused primarily on results and literature generated by Ahmed and colleagues, the, escalation of drug intake with the long-access model has also been observed in many other laboratories (S H Ahmed, 2012; Serge H Ahmed, 2011). We should however mention that there are also experiments that did not report escalation of intake with this model. Yet, as far as we can judge from the published literature, they represent a small minority of cases (reviewed in (Serge H Ahmed, 2011)).

It is also important to note that some of the changes in drug-taking or -seeking behavior seen after extended access to the drug are not necessarily specific. For instance, increased reinstatement of cocaine seeking has also been observed after a short access to the drug in rats trained under a high fixed-ratio schedule of reinforcement (Keiflin, Isingrini, & Cador, 2008). However, the fact that different factors produce similar behavioral outcomes does not rule out distinct underlying neurobiological mechanisms.

Habits or Plans?

One of the most critical debates about the motivational nature of addiction has been whether drug seeking in addicts is a goal-directed behavior, or is it a set of compulsive behaviors that arises due to the dominance of habits. In other words, do addicts search out drugs deliberatively and explicitly for their effects, or are they driven to the drug choices implicitly by the over-blown motivational values? This study suggests that many critical aspects of drug addiction can be understood as a disorder in the brain goal-directed associative learning system that aims at defending the physiological stability of the organism. In other words, drug-seeking can be depicted as planning for fulfilling the potentially escalating need for cocaine. We showed that many aspects of addiction like compulsive drug use, that has been classically attributed to dominance of a habit system (Everitt & Robbins, 2005; Redish, 2004), could be explained by a goal-directed planning system integrated with a homeostatic regulation mechanism.

Critically, habit-based theories of addiction leave the robust pattern of drug self-administration (Serge H Ahmed & Koob, 1998) (initial loading and pause phases, and the forthcoming regular responding) unexplained. This is primarily since habits, by definition, are inflexible to moment-to-moment changes in internal or external conditions. For example, one might define cocaine reward as the expected deviation-reduction of the cocaine-related homeostatic variable (as in our model) but then use a habitual, rather than a goal-directed system for learning to seek such reward. Such a system, in clear contrast to the regular rate of self-administration observed in rats (Serge H Ahmed & Koob, 1998), would always fluctuate between periods of burst and periods of silence of drug-seeking response. This is because when the internal state is below the setpoint, the system learns to attribute a positive habitual value to the self-administration response and keeps responding even when the internal state reaches and overshoots the setpoint. When the habitual system eventually learns that cocaine outcome has become punishing (due to the internal state being above the setpoint), a period of silence starts and continues even after the internal state has dropped far below the setpoint; and so on. The steady pattern of self-administration in rats, however, remains regular even after escalation, suggesting that self-administration remains goal-directed after long-term LgA condition.

Last but not least, habit-based approaches cannot explain the fact that getting access to, procuring and taking drugs in the real world often require complex planning-based behavioral strategies (DiClemente, 2006; Faupel, Weaver, & Corzine, 2009). Devising such complex behavioral strategies is a clear indication of a goal-directed system being in control.

In a nutshell, without arguing against the importance of habits in understanding many aspects of addiction, our results suggest that several key drug-related behavioral patterns emerge from a goal-directed planning system integrated with a homeostatic regulation mechanism.

Homeostatic Regulation-based Theories

As another class of models of addiction, homeostatic regulation-based models successfully explain the regular pattern of drug self-administration (SA) in animal models (Serge H Ahmed & Koob, 1998, 2005; Pickens & Thompson, 1968), attributing it to the animal defending its homeostasis by regularly compensating for the depleted drug level in the brain. However, lack of associative learning systems fundamentally limits the explanatory power of those theories. Due to this shortcoming, they cannot explain behavioral patterns were a learning component is clearly involved. Beyond the learned initiation of drug self-administration in the first place, these include extinction, relapse (Fig. 6), and rapid resumption of drug seeking in LgA, but not ShA animals after removal of the drug-paired punishment (Fig. 5b). Furthermore, those homeostatic-regulation (as opposed to our homeostatic reinforcement learning) accounts are inconsistent with our new data showing a transitory burst of infusion rate after dose reduction (Fig. 11).

Initial homeostatic deviation and environmental deprivation

According to our model, initiation of cocaine use explicitly requires the drug naïve animal to be in a preexisting state of homeostatic deviation before any prior cocaine experience. Indeed, without this initial deficit, the drug value would be zero, and, thus, there would be no initial motivation to take the drug and, *a fortiori*, no risk to develop drug intake escalation under extended drug access conditions. In other words, our model predicts that cocaine should not be inherently reinforcing in every individual but only in those individuals who present a preexisting deficit that can be alleviated by drug use. This assumption is generally consistent with what we know about individual susceptibility to the initial reinforcing effects of drugs of abuse including cocaine (de Wit & Phillips, 2012), and individual vulnerability to cocaine addiction in people (J. C. Anthony, Warner, & Kessler, 1994; J. Anthony et al., 2005). Our model should thus be considered a model of transition to addiction in individuals who are vulnerable to drug use and addiction, as opposed to a model of the resilient majority.

However, since in laboratory conditions the large majority of drug naïve animals learn to self-administer cocaine, provided that the dose is sufficiently high (Carroll & Lac, 1997), our model seems also to imply that in experimental animals, a prior homeostatic deviation would be the norm rather than the exception, as it is in people. A critical question is whether this is a valid implication, and what factors could cause the large majority of experimental animals to be in a preexisting state of homeostatic deviation that renders them sensitive to the rewarding effects of cocaine.

At present, it is hard to address this question with certainty, mainly because there is currently no objective measure of the homeostatic deviation other than the drug self-administration behavior itself. As our neurobiological framework suggests, we suspect that this deviation is related to a preexisting hypoactivity in NAc D1 neuronal activity (see subsection “Neural substrates”) but this is difficult to establish in the absence of a reference population of animals entirely immune to cocaine self-administration. So we can only speculate about the potential causes of this postulated initial deviation. One highly plausible explanation is that this deviation could be caused by insufficient environmental stimulation of the brain reward system of laboratory animals (Serge H Ahmed & Koob, 2005; Serge H Ahmed, 2005). The environment in which laboratory animals are raised and tested for cocaine self-administration differs indeed considerably from their natural ecological niche in many respects. Notably, it lacks many types of important environmental and behavioral stimulation, including, for instance, sensory, affective, emotional, social and sexual stimulations (Serge H Ahmed, 2005; Bruce K. Alexander & Hadaway, 1982; Heilig, Epstein, Nader, & Shaham, 2016). Such artificial conditions should induce in most exposed individuals many deviations from the species-normative behavioral and need states, including the state of the brain reward system itself. This would explain why the large majority of naïve animals experience cocaine as reinforcing and, thus, are at risk to escalate cocaine use under extended drug access conditions. One unique prediction from this hypothesis is that making the environment more ecological for the nonhuman species under study, notably by increasing and/or diversifying access to nondrug options should reduce cocaine use and thus the risk of drug escalation, at least in the majority of individuals.

In fact, many studies already support this hypothesis. One direct way to test this hypothesis consists of reconstructing in the laboratory a small version of the real world, as was done in the famous Rat Park experiment (B K Alexander, Beyerstein, Hadaway, & Coambs, 1981; B

K Alexander & Hadaway, 1982). In Rat Park, rats lived together in a large enriched colony that offered several different behavioral options, including taking morphine orally from a drinking bottle. When compared to rats living alone in a standard housing cage, rats living in Rat Park drank much less morphine. This outcome shows that the possibility to engage in other pursuits during drug access, including social interactions, can prevent drug use in most rats. Over the past 30 years, this seminal study has been generally reproduced under several different conditions (Bardo, Klebaur, Valone, & Deaton, 2001; Carroll, Lac, & Nygaard, 1989; Carroll & Lac, 1993; Green, Gehrke, & Bardo, 2002; Howes, Dalley, Morrison, Robbins, & Everitt, 2000; M A Nader & Woolverton, 1991; M.A. Nader & Woolverton, 1992; Negus, 2003; Schenk, Lacelle, Gorman, & Amit, 1987). For instance, these environmental enrichments include providing rats with novel objects and activities (Bardo et al., 2001; Green et al., 2002), access to non-drug behavioral alternatives (Carroll et al., 1989; Carroll & Lac, 1993; Lenoir & Ahmed, 2008; Lenoir, Serre, Cantin, & Ahmed, 2007), and opportunity for social interaction (Heilig et al., 2016; Howes et al., 2000).

In humans, preexisting hedonic, affective and/or motivational need states are thought to result from complex interactions between genes, developmental history and the current situation of the individual (B K Alexander & Hadaway, 1982; Altman et al., 1996; Badiani, Belin, Epstein, Calu, & Shaham, 2011; Glantz & Pickens RW, 1992; Heilig et al., 2016; Higgins, Heil, & Lussier, 2004; Kendler et al., 2012; Koob & Le Moal, 2001). In this respect, laboratory animals probably represent more of a model of the human population at risk to develop cocaine use and addiction than a model of the general population, as commonly assumed (Serge H Ahmed & Koob, 2005; Serge H Ahmed, 2005). This parallelism provides a foundation that can allow us to extrapolate, at least to some extent, the conclusions and predictions of our theoretical model from laboratory animals to people who are vulnerable to addiction.

Relevance to Human Addiction

There is little quantitative data on escalated cocaine use in humans comparable to that obtainable in laboratory animals. This explains why we only focused here on animal data in the present study. However, it is important to stress that most, if not all, of the qualitative conclusions from our model correspond well with what we know about the transition to escalated cocaine use in humans. The most important correspondences include: i) regulation or self-titration of drug intake to the dose available (Lynch et al., 2006; Sughondhabirom et al., 2005); ii) escalation of drug intake over time (Hser, Evans, Huang, Brecht, & Li, 2008; Shaffer & Eber, 2002; Siegel, 1984); iii) post-escalation increase in drug-primed craving (Jaffe, Cascella, Kumor, & Sherer, 1989; Nora D Volkow et al., 2006); and, finally, iv) continued predominance of controlled, goal-directed drug-seeking and drug-talking behaviors in people with addiction (Baumeister & Vonasch, 2015; Hogarth & Troisi, 2015; Pickard, 2012). Furthermore, several predictions of our model are also consistent with some well-entrenched clinical practices. For instance, the prediction that escalated levels of drug use can be reversed only when access to the drug is sufficiently decreased is consistent and even supports recommendations of total abstinence or of low levels of drug use to people with addiction. Our model also predicts that the well-known priming effect of cocaine on craving should be extinguishable (Mahoney, Kalechstein, De La Garza, & Newton, 2007). As far as we can tell, this prediction is unique to our model and could be tested in humans using cocaine under medical supervision or an analogue of cocaine that does not cross the blood brain barrier (Roy A Wise, Wang, & You, 2008). However, as our model predicts, the extinction of drug craving with repeated drug priming should not be expected to have protective effects against relapse, a conclusion that is generally consistent with other cue extinction procedures (Conklin & Tiffany, 2002). Finally, the notion that drug seeking and taking largely remains goal-directed, even after escalation, may help to explain why after some initial relapses, people with addiction nevertheless remain able, at some later stages of their life, to quit drugs when having the opportunity to engage in other goals or pursuits that are incompatible with continued drug use (Heyman, 2013).

Funding and Disclosure:

All authors of this study have no conflict of interest. M.K. and B.G. are supported by grants from Frontiers du Vivant, the French MESR, CNRS, INSERM, ANR, ENP and NERF. S.H.A. is supported by the French Research Council (CNRS), the Université de Bordeaux, the French National Agency (ANR, ANR2010-BLAN-1404-01), the Conseil Régional d'Aquitaine (CRA20101301022; CRA11004375/11004699) and the LabEX BRAIN (ANR-10-LABX-43). BG also acknowledges support from the Russian Academic Excellence Project '5-100'.

Acknowledgments:

M.K., B.G. and S.H.A. defined the project. M.K. designed the model, performed the simulations and analyzed experimental data. P.G. and A.D. performed the experiment. S.H.A. provided the raw data from previous experiments. M.K., S.H.A. and B.G. discussed the results. M.K. wrote the paper with the help of S.H.A. and B.G. All authors critically reviewed content and approved final version for publication. We thank Peter Dayan for critical discussions.

References

Ahmed, S. H. (2005). Imbalance between drug and non-drug reward availability: a major risk factor for addiction. *European Journal of Pharmacology*, *526*(1-3), 9–20.

Ahmed, S. H. (2011). Toward an evolutionary basis for resilience to drug addiction. *The Behavioral and Brain Sciences*, *34*(6), 310–1.

Ahmed, S. H. (2012). The science of making drug-addicted animals. *Neuroscience*, *211*, 107–25.

Ahmed, S. H., & Cador, M. (2006). Dissociation of psychomotor sensitization from compulsive cocaine consumption. *Neuropsychopharmacology*, *31*(3), 563–71.

Ahmed, S. H., & Koob, G. F. (1998). Transition from moderate to excessive drug intake: change in hedonic set point. *Science*, *282*(5387), 298–300.

Ahmed, S. H., & Koob, G. F. (1999). Long-lasting increase in the set point for cocaine self-administration after escalation in rats. *Psychopharmacology*, *146*(3), 303–12.

Ahmed, S. H., & Koob, G. F. (2004). Changes in response to a dopamine receptor antagonist in rats with escalating cocaine intake. *Psychopharmacology*, *172*(4), 450–4.

Ahmed, S. H., & Koob, G. F. (2005). Transition to drug addiction: a negative reinforcement model based on an allostatic decrease in reward function. *Psychopharmacology*, *180*(3), 473–490.

Ahmed, S. H., Lin, D., Koob, G. F., & Parsons, L. H. (2003). Escalation of cocaine self-administration does not depend on altered cocaine-induced nucleus accumbens dopamine levels. *Journal of Neurochemistry*, *86*(1), 102–13.

Alexander, B. K., Beyerstein, B. L., Hadaway, P. F., & Coambs, R. B. (1981). Effect of early and later colony housing on oral ingestion of morphine in rats. *Pharmacology, Biochemistry, and Behavior*, *15*(4), 571–6.

Alexander, B. K., & Hadaway, P. F. (1982). Opiate addiction: The case for an adaptive orientation. *Psychological Bulletin*, *92*(2), 367–381.

Altman, J., Everitt, B. J., Glautier, S., Markou, A., Nutt, D., Oretti, R., … Robbins, T. W. (1996). The biological, social and clinical bases of drug addiction: commentary and debate. *Psychopharmacology*, *125*(4), 285–345.

Anthony, J. C., Warner, L. A., & Kessler, R. C. (1994). Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. *Experimental and Clinical Psychopharmacology*, *2*(3), 244{\textendash}268. http://doi.org/10.1037/1064-1297.2.3.244

Anthony, J., Chen, C., Storr, C., Hughes, M., Anthony, J. C., & Nelson, C. B. (2005). Drug dependence epidemiology. *Clinical Neuroscience Research*, *5*, 55–68.

Bachtell, R. K., Whisler, K., Karanian, D., & Self, D. W. (2005). Effects of intra-nucleus accumbens shell administration of dopamine agonists and antagonists on cocaine-taking and cocaine-seeking behaviors in the rat. *Psychopharmacology*, *183*(1), 41–53.

Badiani, A., Belin, D., Epstein, D., Calu, D., & Shaham, Y. (2011). Opiate versus psychostimulant addiction: the differences do matter. *Nature Reviews. Neuroscience*, *12*(11), 685–700.

Baker, D. A., McFarland, K., Lake, R. W., Shen, H., Tang, X.-C., Toda, S., & Kalivas, P. W. (2003). Neuroadaptations in cystine-glutamate exchange underlie cocaine relapse.

Nature Neuroscience, *6*(7), 743–9.

Bardo, M. T., Klebaur, J. E., Valone, J. M., & Deaton, C. (2001). Environmental enrichment decreases intravenous self-administration of amphetamine in female and male rats. *Psychopharmacology*, *155*(3), 278–84.

Bari, A. A., & Pierce, R. C. (2005). D1-like and D2 dopamine receptor antagonists administered into the shell subregion of the rat nucleus accumbens decrease cocaine, but not food, reinforcement. *Neuroscience*, *135*(3), 959–968.

Baumeister, R. F., & Vonasch, A. J. (2015). Uses of self-regulation to facilitate and restrain addictive behavior. *Addictive Behaviors*, *44*, 3–8.

Ben-Shahar, O., Keeley, P., Cook, M., Brake, W., Joyce, M., Nyffeler, M., … Ettenberg, A. (2007). Changes in levels of D1, D2, or NMDA receptors during withdrawal from brief or extended daily access to IV cocaine. *Brain Research*, *1131*, 220–228.

Ben-Shahar, O. M., Szumlinski, K. K., Lominac, K. D., Cohen, A., Gordon, E., Ploense, K. L., … Woodward, N. (2012). Extended access to cocaine self-administration results in reduced glutamate function within the medial prefrontal cortex. *Addiction Biology*, *17*(4), 746–57.

Bertran-Gonzalez, J., Laurent, V., Chieng, B. C., Christie, M. J., & Balleine, B. W. (2013). Learning-related translocation of δ-opioid receptors on ventral striatal cholinergic interneurons mediates choice between goal-directed actions. *The Journal of Neuroscience*, *33*(41), 16060–71.

Bock, R., Shin, J. H., Kaplan, A. R., Dobi, A., Markey, E., Kramer, P. F., … Alvarez, V. A. (2013). Strengthening the accumbal indirect pathway promotes resilience to compulsive cocaine use. *Nature Neuroscience*, *16*(5), 632–8.

Briand, L. A., Flagel, S. B., Garcia-Fuster, M. J., Watson, S. J., Akil, H., Sarter, M., & Robinson, T. E. (2008). Persistent alterations in cognitive function and prefrontal dopamine D2 receptors following extended, but not limited, access to self-administered cocaine. *Neuropsychopharmacology*, *33*(12), 2969–80.

Caine, S. B., Heinrichs, S. C., Coffin, V. L., & Koob, G. F. (1995). Effects of the dopamine D-1 antagonist SCH 23390 microinjected into the accumbens, amygdala or striatum on cocaine self-administration in the rat. *Brain Research*, *692*(1-2), 47–56.

Calipari, E. S., Bagot, R. C., Purushothaman, I., Davidson, T. J., Yorgason, J. T., Peña, C. J., … Nestler, E. J. (2016). In vivo imaging identifies temporal signature of D1 and D2 medium spiny neurons in cocaine reward. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2726–31.

Carlezon, W. A., & Thomas, M. J. (2009). Biological substrates of reward and aversion: A nucleus accumbens activity hypothesis. *Neuropharmacology*, *56*, 122–132.

Carroll, M. E., & Lac, S. T. (1993). Autoshaping i.v. cocaine self-administration in rats: effects of nondrug alternative reinforcers on acquisition. *Psychopharmacology*, *110*(1-2), 5–12.

Carroll, M. E., & Lac, S. T. (1997). Acquisition of i.v. amphetamine and cocaine self-administration in rats as a function of dose. *Psychopharmacology*, *129*(3), 206–14.

Carroll, M. E., Lac, S. T., & Nygaard, S. L. (1989). A concurrently available nondrug reinforcer prevents the acquisition or decreases the maintenance of cocaine-reinforced behavior. *Psychopharmacology*, *97*(1), 23–9.

Chandra, R., Lenz, J. D., Gancarz, A. M., Chaudhury, D., Schroeder, G. L., Han, M.-H., …

Lobo, M. K. (2013). Optogenetic inhibition of D1R containing nucleus accumbens neurons alters cocaine-mediated regulation of Tiam1. *Frontiers in Molecular Neuroscience*, *6*, 13.

Conklin, C. A., & Tiffany, S. T. (2002). Applying extinction research and theory to cue-exposure addiction treatments. *Addiction (Abingdon, England)*, *97*(2), 155–67.

Conrad, K. L., Ford, K., Marinelli, M., & Wolf, M. E. (2010). Dopamine receptor expression and distribution dynamically change in the rat nucleus accumbens after withdrawal from cocaine self-administration. *Neuroscience*, *169*(1), 182–194.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711.

de Wit, H., & Phillips, T. J. (2012). Do initial responses to drugs predict future use or abuse? *Neuroscience and Biobehavioral Reviews*, *36*(6), 1565–76.

Desai, R. I., Paronis, C. A., Martin, J., Desai, R., & Bergman, J. (2010). Monoaminergic psychomotor stimulants: discriminative stimulus effects and dopamine efflux. *The Journal of Pharmacology and Experimental Therapeutics*, *333*(3), 834–43.

Dezfouli, A., Piray, P., Keramati, M. M., Ekhtiari, H., Lucas, C., & Mokri, A. (2009). A neurocomputational model for cocaine addiction. *Neural Computation*, *21*(10), 2869–2893.

Di Chiara, G., & Imperato, A. (1988). Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(14), 5274–5278.

DiClemente, C. C. (2006). *Addiction and Change: How Addictions Develop and Addicted People Recover*. The Guilford Press.

Dobbs, L. K., Kaplan, A. R., Lemos, J. C., Matsui, A., Rubinstein, M., & Alvarez, V. A. (2016). Dopamine Regulation of Lateral Inhibition between Striatal Neurons Gates the Stimulant Actions of Cocaine. *Neuron*, *90*(5), 1100–1113.

Doyle, S. E., Ramôa, C., Garber, G., Newman, J., Toor, Z., & Lynch, W. J. (2014). A Shift in the Role of Glutamatergic Signaling in the Nucleus Accumbens Core With the Development of an Addicted Phenotype. *Biological Psychiatry*, *76*(10), 810–815.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–1489.

Faupel, C. E., Weaver, G. S., & Corzine, J. (2009). *The Sociology of American Drug Use*. (O. U. Press, Ed.) (3rd Editio).

Frantz, K. J., O'Dell, L. E., & Parsons, L. H. (2007). Behavioral and neurochemical responses to cocaine in periadolescent and adult rats. *Neuropsychopharmacology*, *32*(3), 625–37.

Gerfen, C. R., & Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine. *Annual Review of Neuroscience*, *34*, 441–66.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197–209.

Glantz, M., & Pickens RW. (1992). *Vulnerability to drug abuse*. Washington, DC: American Psychological Association. Retrieved from https://www.google.co.uk/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=Vulnerability+to+drug+abuse.+American+Psychological+Association

Green, T. A., Gehrke, B. J., & Bardo, M. T. (2002). Environmental enrichment decreases intravenous amphetamine self-administration in rats: dose-response functions for fixed- and progressive-ratio schedules. *Psychopharmacology*, *162*(4), 373–8.

Guillem, K., Ahmed, S. H., & Peoples, L. L. (2014). Escalation of Cocaine Intake and Incubation of Cocaine Seeking Are Correlated with Dissociable Neuronal Processes in Different Accumbens Subregions. *Biological Psychiatry*, *76*(1), 31–39.

Hao, Y., Martin-Fardon, R., & Weiss, F. (2010). Behavioral and functional evidence of metabotropic glutamate receptor 2/3 and metabotropic glutamate receptor 5 dysregulation in cocaine-escalated rats: factor in the transition to dependence. *Biological Psychiatry*, *68*(3), 240–8.

Heilig, M., Epstein, D. H., Nader, M. A., & Shaham, Y. (2016). Time to connect: bringing social context into addiction neuroscience. *Nature Reviews Neuroscience*, *17*(9), 592–599.

Heyman, G. M. (2013). Quitting drugs: quantitative and qualitative features. *Annual Review of Clinical Psychology*, *9*, 29–59.

Higgins, S. T., Heil, S. H., & Lussier, J. P. (2004). Clinical implications of reinforcement as a determinant of substance use disorders. *Annual Review of Psychology*, *55*, 431–61.

Hikida, T., Kimura, K., Wada, N., Funabiki, K., & Nakanishi, S. (2010). Distinct Roles of Synaptic Transmission in Direct and Indirect Striatal Pathways to Reward and Aversive Behavior. *Neuron*, *66*(6), 896–907.

Hogarth, L., & Troisi, J. R. (2015). A hierarchical instrumental decision theory of nicotine dependence. *Current Topics in Behavioral Neurosciences*, *23*, 165–91.

Howes, S. R., Dalley, J. W., Morrison, C. H., Robbins, T. W., & Everitt, B. J. (2000). Leftward shift in the acquisition of cocaine self-administration in isolation-reared rats: relationship to extracellular levels of dopamine, serotonin and glutamate in the nucleus accumbens and amygdala-striatal FOS expression. *Psychopharmacology*, *151*(1), 55–63.

Hser, Y.-I., Evans, E., Huang, D., Brecht, M.-L., & Li, L. (2008). Comparing the dynamic course of heroin, cocaine, and methamphetamine use over 10 years. *Addictive Behaviors*, *33*(12), 1581–1589.

Jaffe, J. H., Cascella, N. G., Kumor, K. M., & Sherer, M. A. (1989). Cocaine-induced cocaine craving. *Psychopharmacology*, *97*(1), 59–64.

Kalivas, P. W. (2009). The glutamate homeostasis hypothesis of addiction. *Nature Reviews. Neuroscience*, *10*(8), 561–572.

Keiflin, R., Isingrini, E., & Cador, M. (2008). Cocaine-induced reinstatement in rats: evidence for a critical role of cocaine stimulus properties. *Psychopharmacology*, *197*(4), 649–60.

Kendler, K. S., Chen, X., Dick, D., Maes, H., Gillespie, N., Neale, M. C., & Riley, B. (2012). Recent advances in the genetic epidemiology and molecular genetics of substance use disorders. *Nature Neuroscience*, *15*(2), 181–9.

Keramati, M., & Gutkin, B. (2013). Imbalanced Decision Hierarchy in Addicts Emerging from Drug-Hijacked Dopamine Spiraling Circuit. *PLoS ONE*, *8*(4), e61489.

Keramati, M., & Gutkin, B. (2014). Homeostatic reinforcement learning for integrating reward collection and physiological stability. *eLife*, *3*.

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, *24*(2), 97–129.

Kourrich, S., Calu, D. J., & Bonci, A. (2015). Intrinsic plasticity: an emerging player in addiction. *Nature Reviews. Neuroscience*, *16*(3), 173–84.

Lamas, X., Negus, S. S., Gatch, M. B., & Mello, N. K. (1998). Effects of heroin/cocaine combinations in rats trained to discriminate heroin or cocaine from saline. *Pharmacology, Biochemistry, and Behavior*, *60*(2), 357–64.

Lenoir, M., & Ahmed, S. H. (2008). Supply of a nondrug substitute reduces escalated heroin consumption. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *33*(9), 2272–2282. http://doi.org/10.1038/sj.npp.1301602

Lenoir, M., Serre, F., Cantin, L., & Ahmed, S. H. (2007). Intense sweetness surpasses cocaine reward. *PloS One*, *2*(8), e698.

Lobo, M. K., Covington, H. E., Chaudhury, D., Friedman, A. K., Sun, H., Damez-Werno, D., … Nestler, E. J. (2010). Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science*, *330*(6002), 385–90.

Lobo, M. K., & Nestler, E. J. (2011). The striatal balancing act in drug addiction: distinct roles of direct and indirect pathway medium spiny neurons. *Frontiers in Neuroanatomy*, *5*, 41.

Luo, Z., Volkow, N. D., Heintz, N., Pan, Y., & Du, C. (2011). Acute cocaine induces fast activation of D1 receptor and progressive deactivation of D2 receptor striatal neurons: in vivo optical microprobe [Ca2+]i imaging. *The Journal of Neuroscience*, *31*(37), 13180–90.

Lüscher, C., & Malenka, R. C. (2011). Drug-Evoked Synaptic Plasticity in Addiction: From Molecular Changes to Circuit Remodeling. *Neuron*, *69*(4), 650–663.

Lynch, W. J., Sughondhabirom, A., Pittman, B., Gueorguieva, R., Kalayasiri, R., Joshua, D., … Malison, R. T. (2006). A paradigm to investigate the regulation of cocaine self-administration in human cocaine users: a randomized trial. *Psychopharmacology*, *185*(3), 306–14.

Mahoney, J. J., Kalechstein, A. D., De La Garza, R., & Newton, T. F. (2007). A qualitative and quantitative review of cocaine-induced craving: The phenomenon of priming. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *31*(3), 593–599.

Mantsch, J. R., & Goeders, N. E. (1998). Generalization of a restraint-induced discriminative stimulus to cocaine in rats. *Psychopharmacology*, *135*(4), 423–6.

Martinez, D., Slifstein, M., Narendran, R., Foltin, R. W., Broft, A., Hwang, D.-R., … Laruelle, M. (2009). Dopamine D1 receptors in cocaine dependence measured with PET and the choice to self-administer cocaine. *Neuropsychopharmacology*, *34*(7), 1774–1782.

Mihindou, C., Vouillac, C., Koob, G. F., & Ahmed, S. H. (2011). Preclinical validation of a novel cocaine exposure therapy for relapse prevention. *Biological Psychiatry*, *70*(6), 593–8.

Nader, M. A., Czoty, P. W., Gould, R. W., & Riddick, N. V. (2008). Positron emission tomography imaging studies of dopamine receptors in primate models of addiction. *Philosophical Transactions of the Royal Society of London*, *363*(1507), 3223–3232.

Nader, M. A., Daunais, J. B., Moore, T., Nader, S. H., Moore, R. J., Smith, H. R., … Porrino, L. J. (2002). Effects of cocaine self-administration on striatal dopamine systems in rhesus monkeys: initial and chronic exposure. *Neuropsychopharmacology*, *27*(1), 35–46.

Nader, M. A., Morgan, D., Gage, H. D., Nader, S. H., Calhoun, T. L., Buchheimer, N., … Mach, R. H. (2006). PET imaging of dopamine D2 receptors during chronic cocaine self-administration in monkeys. *Nature Neuroscience*, *9*(8), 1050–1056.

Nader, M. A., Morgan, D., Grant, K. A., Gage, H. D., Mach, R. H., Kaplan, J. R., … Ehrenkaufer, R. L. (2002). Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration. *Nature Neuroscience*, *5*(2), 169–174.

Nader, M. A., & Woolverton, W. L. (1991). Effects of increasing the magnitude of an alternative reinforcer on drug choice in a discrete-trials choice procedure. *Psychopharmacology*, *105*(2), 169–74.

Nader, M. A., & Woolverton, W. L. (1992). Choice between cocaine and food by rhesus monkeys: effects of conditions of food availability. *Behavioural Pharmacology*, *3*(6), 635–638.

Negus, S. S. (2003). Rapid assessment of choice between cocaine and food in rhesus monkeys: effects of environmental manipulations and treatment with d-amphetamine and flupenthixol. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *28*(5), 919–31.

Newlin, D. B., Regalia, P. A., Seidman, T. I., & Bobashev, G. (2012). Control Theory and Addictive Behavior. In B. Gutkin & S. H. Ahmed (Eds.), *Computational Neuroscience of Drug Addiction* (pp. 57–108). New York, NY: Springer New York.

Park, K., Volkow, N. D., Pan, Y., & Du, C. (2013). Chronic cocaine dampens dopamine signaling during cocaine intoxication and unbalances D1 over D2 receptor signaling. *The Journal of Neuroscience*, *33*(40), 15827–36.

Pascoli, V., Terrier, J., Espallergues, J., Valjent, E., O'Connor, E. C., & Lüscher, C. (2014). Contrasting forms of cocaine-evoked plasticity control components of relapse. *Nature*, *509*(7501), 459–64.

Peoples, L. L., & Cavanaugh, D. (2003). Differential changes in signal and background firing of accumbal neurons during cocaine self-administration. *Journal of Neurophysiology*, *90*(2), 993–1010.

Peoples, L. L., Gee, F., Bibi, R., & West, M. O. (1998). Phasic firing time locked to cocaine self-infusion and locomotion: dissociable firing patterns of single nucleus accumbens neurons in the rat. *The Journal of Neuroscience*, *18*(18), 7588–98.

Peoples, L. L., Kravitz, A. V, Lynch, K. G., & Cavanaugh, D. J. (2007). Accumbal neurons that are activated during cocaine self-administration are spared from inhibitory effects of repeated cocaine self-administration. *Neuropsychopharmacology*, *32*(5), 1141–58.

Peoples, L. L., Kravitz, A. V., & Karine, G. (2010). *Application of Chronic Extracellular Recording Method to Studies of Cocaine Self-Administration: Method and Progress. Advances in the Neuroscience of Addiction*.

Peoples, L. L., Uzwiak, A. J., Gee, F., & West, M. O. (1999). Tonic firing of rat nucleus accumbens neurons: changes during the first 2 weeks of daily cocaine self-administration sessions. *Brain Research*, *822*(1), 231–236.

Pettit, H. O., & Justice, J. B. (1989). Dopamine in the nucleus accumbens during cocaine self-administration as studied by in vivo microdialysis. *Pharmacology, Biochemistry, and Behavior*, *34*(4), 899–904.

Piazza, P. V, Deroche-Gamonent, V., Rouge-Pont, F., & Le Moal, M. (2000). Vertical shifts in self-administration dose-response functions predict a drug-vulnerable phenotype predisposed to addiction. *The Journal of Neuroscience*, *20*(11), 4226–32.

Pickard, H. (2012). The Purpose in Chronic Addiction. *AJOB Neuroscience*, *3*(2), 40–49.

Pickens, R., & Thompson, T. (1968). Cocaine-reinforced behavior in rats: effects of reinforcement magnitude and fixed-ratio size. *The Journal of Pharmacology and Experimental Therapeutics*, *161*(1), 122–9.

Rangel, A. (2013). Regulation of dietary choice by the decision-making circuitry. *Nature Neuroscience*, *16*(12), 1717–24.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews. Neuroscience*, *9*(7), 545–56.

Redish, A. D. (2004). Addiction as a computational process gone awry. *Science*, *306*(5703), 1944–1947.

Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychological Review*, *114*(3), 784–805.

Root, D. H., Melendez, R. I., Zaborszky, L., & Napier, T. C. (2015). The ventral pallidum: Subregion-specific functional anatomy and roles in motivated behaviors. *Progress in Neurobiology*, *130*, 29–70.

Schenk, S., Lacelle, G., Gorman, K., & Amit, Z. (1987). Cocaine self-administration in rats influenced by environmental conditions: implications for the etiology of drug abuse. *Neuroscience Letters*, *81*(1-2), 227–31.

Scofield, M. D., Heinsbroek, J. A., Gipson, C. D., Kupchik, Y. M., Spencer, S., Smith, A. C. W., … Kalivas, P. W. (2016). The Nucleus Accumbens: Mechanisms of Addiction across Drug Classes Reflect the Importance of Glutamate Homeostasis. *Pharmacological Reviews*, *68*(3), 816–71.

Self, D. W. (2014). Diminished Role for Dopamine D1 Receptors in Cocaine Addiction? *Biological Psychiatry*.

Shaffer, H. J., & Eber, G. B. (2002). Temporal progression of cocaine dependence symptoms in the US National Comorbidity Survey. *Addiction (Abingdon, England)*, *97*(5), 543–54.

Siegel, R. K. (1984). Changing patterns of cocaine use: longitudinal observations, consequences, and treatment. *NIDA Research Monograph*, *50*, 92–110.

Smith, K. S., Tindell, A. J., Aldridge, J. W., & Berridge, K. C. (2009). Ventral pallidum roles in reward and motivation. *Behavioural Brain Research*, *196*(2), 155–167.

Smith, R. J., Lobo, M. K., Spencer, S., & Kalivas, P. W. (2013). Cocaine-induced adaptations in D1 and D2 accumbens projection neurons (a dichotomy not necessarily synonymous with direct and indirect pathways). *Current Opinion in Neurobiology*, *23*(4), 546–552.

Soares-Cunha, C., Coimbra, B., Sousa, N., & Rodrigues, A. J. (2016). Reappraising striatal D1- and D2-neurons in reward and aversion. *Neuroscience & Biobehavioral Reviews*, *68*, 370–386.

Sughondhabirom, A., Jain, D., Gueorguieva, R., Coric, V., Berman, R., Lynch, W. J., … Malison, R. T. (2005). A paradigm to investigate the self-regulation of cocaine administration in humans. *Psychopharmacology*, *180*(3), 436–46.

Suto, N., Ecke, L. E., & Wise, R. A. (2009). Control of within-binge cocaine-seeking by dopamine and glutamate in the core of nucleus accumbens. *Psychopharmacology*, *205*(3), 431–9.

Suto, N., & Wise, R. A. (2011). Satiating effects of cocaine are controlled by dopamine actions in the nucleus accumbens core. *The Journal of Neuroscience*, *31*(49), 17917–22.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.

Terrier, J., Lüscher, C., & Pascoli, V. (2016). Cell-Type Specific Insertion of GluA2-Lacking AMPARs with Cocaine Exposure Leading to Sensitization, Cue-Induced Seeking, and Incubation of Craving. *Neuropsychopharmacology*, *41*(7), 1779–89.

Thanos, P. K., Michaelides, M., Umegaki, H., & Volkow, N. D. (2008). D2R DNA transfer into the nucleus accumbens attenuates cocaine self-administration in rats. *Synapse*, *62*(7), 481–6.

Tsibulsky, V. L., & Norman, A. B. (1999). Satiety threshold: a quantitative model of maintained cocaine self-administration. *Brain Research*, *839*(1), 85–93.

Volkow, N. D., Fowler, J. S., Wang, G. J., Hitzemann, R., Logan, J., Schlyer, D. J., … Wolf, A. P. (1993). Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers. *Synapse*, *14*(2), 169–77.

Volkow, N. D., & Morales, M. (2015). The Brain on Drugs: From Reward to Addiction. *Cell*, *162*(4), 712–725.

Volkow, N. D., Wang, G. J., Fowler, J. S., Logan, J., Gatley, S. J., Hitzemann, R., … Pappas, N. (1997). Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature*, *386*(6627), 830–3.

Volkow, N. D., Wang, G.-J., Fowler, J. S., Logan, J., Gatley, S. J., Gifford, A., … Pappas, N. (1999). Prediction of reinforcing responses to psychostimulants in humans by brain dopamine D2 receptor levels. *The American Journal of Psychiatry*, *156*(9), 1440–1443.

Volkow, N. D., Wang, G.-J., Telang, F., Fowler, J. S., Logan, J., Childress, A.-R., … Wong, C. T. (2006). Cocaine cues and dopamine in dorsal striatum: mechanism of craving in cocaine addiction. *The Journal of Neuroscience*, *26*(24), 6583–6588.

Wee, S., Specio, S. E., & Koob, G. F. (2007). Effects of dose and session duration on cocaine self-administration in rats. *The Journal of Pharmacology and Experimental Therapeutics*, *320*(3), 1134–43.

Wheeler, R. A., & Carelli, R. M. (2009). Dissecting motivational circuitry to understand substance abuse. *Neuropharmacology*, *56*, 149–159.

Wise, R. A., Newton, P., Leeb, K., Burnette, B., Pocock, D., & Justice, J. B. (1995). Fluctuations in nucleus accumbens dopamine concentration during intravenous cocaine self-administration in rats. *Psychopharmacology*, *120*(1), 10–20.

Wise, R. A., Wang, B., & You, Z.-B. (2008). Cocaine serves as a peripheral interoceptive conditioned stimulus for central glutamate and dopamine release. *PloS One*, *3*(8), e2846.

Wolf, M. E. (2016). Synaptic mechanisms underlying persistent cocaine craving. *Nature Reviews. Neuroscience*, *17*(6), 351–65.

Yager, L. M., Garcia, A. F., Wunsch, A. M., & Ferguson, S. M. (2015). The ins and outs of the striatum: Role in drug addiction. *Neuroscience*, *301*, 529–541.

Zahm, D. S. (2000). An integrative neuroanatomical perspective on some subcortical substrates of adaptive responding with emphasis on the nucleus accumbens. *Neuroscience & Biobehavioral Reviews*, *24*(1), 85–105.

Zittel-Lazarini, A., Cador, M., & Ahmed, S. H. (2007). A critical transition in cocaine self-administration: behavioral and neurobiological implications. *Psychopharmacology*, *192*(3), 337–46.

Figure Legends

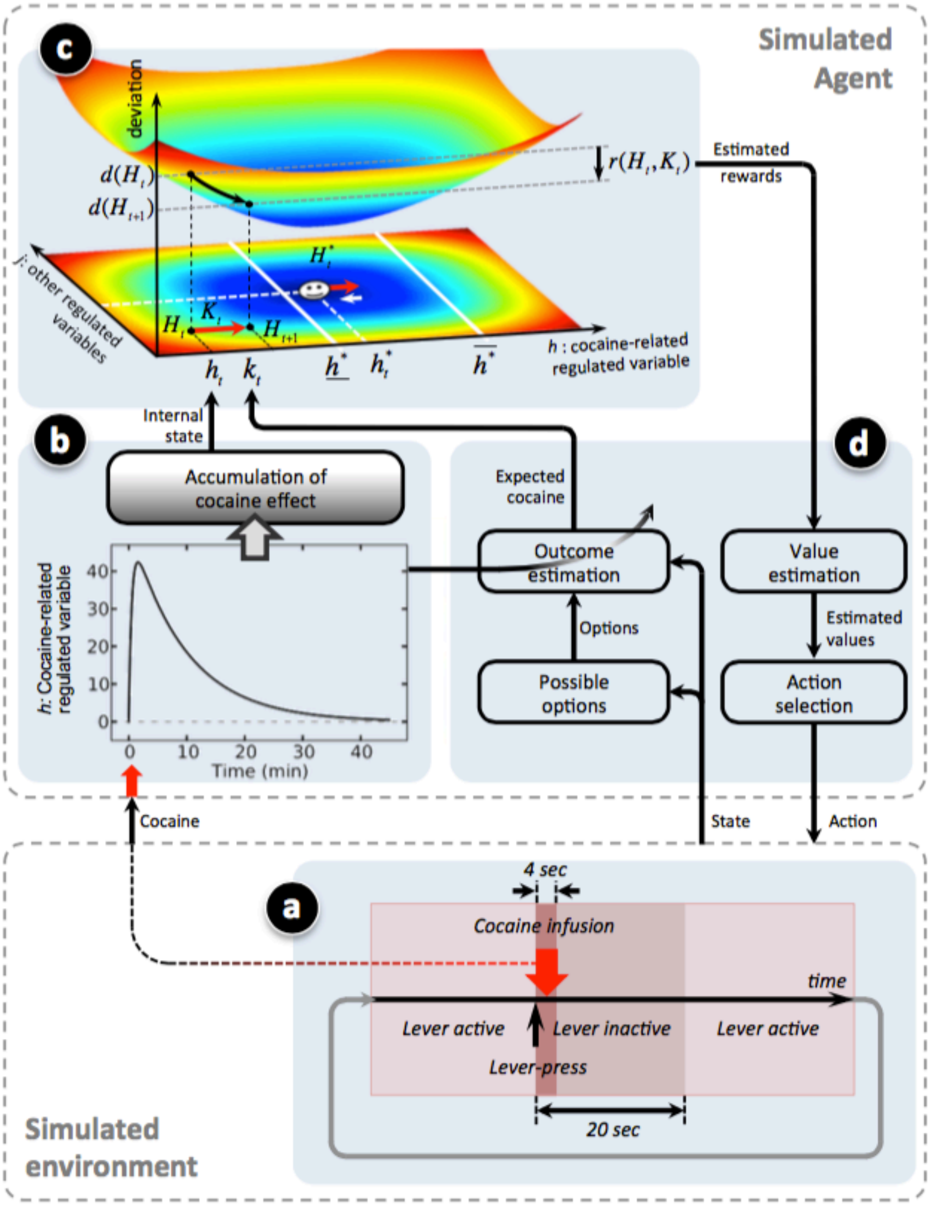

Fig. 1: Schematics of the model. (a) The model is simulated in a self-administration paradigm where each lever-press (fixed-ratio one) initiates an intravenous infusion of cocaine over 4sec, followed by a 20sec time-out period during which the lever is inactive. See (S H Ahmed & Koob, 1999; S H Ahmed, 2012; Serge H Ahmed & Cador, 2006; Serge H Ahmed & Koob, 1998, 2005; Hao et al., 2010; Mihindou et al., 2011; Piazza et al., 2000; Wee et al., 2007; Zittel-Lazarini et al., 2007) (b) A brain internal variable (h) increases initially upon an infusion of the drug and then falls gradually as circulating cocaine degrades. Dynamics of h are compatible with cocaine-induced pharmacodynamics of tonic dopamine in the NAc (Frantz et al., 2007). Effect of consecutive infusions accumulates over time. (c) The rewarding value (r) of an outcome (e.g. a given dose of cocaine, indicated by K_t) is equal to the evoked decrease in the distance (drive, indicated by $d(H_t)$) from the internal state (H_t) to

the homeostatic setpoint (H^*). Note that H_t is a vector composed of h_t and other homeostatically-regulated variables. In parallel with this acute effect, every cocaine infusion triggers a slow adaptive process that shifts the setpoint forward. Absence of cocaine results in slow recovery of the setpoint to its initial level. Note that $\underline{h^*}$ and $\overline{h^*}$ indicate the lower and upper bounds of the setpoint, respectively. (d) Given a current external state, the agent predicts the expected outcomes of possible action choices and based on that, estimates the drive-reduction rewarding values of the choices. According to the estimated values, the agent selects the action. The curved arrow represents updating outcome expectancies based on feedbacks received from the environment.

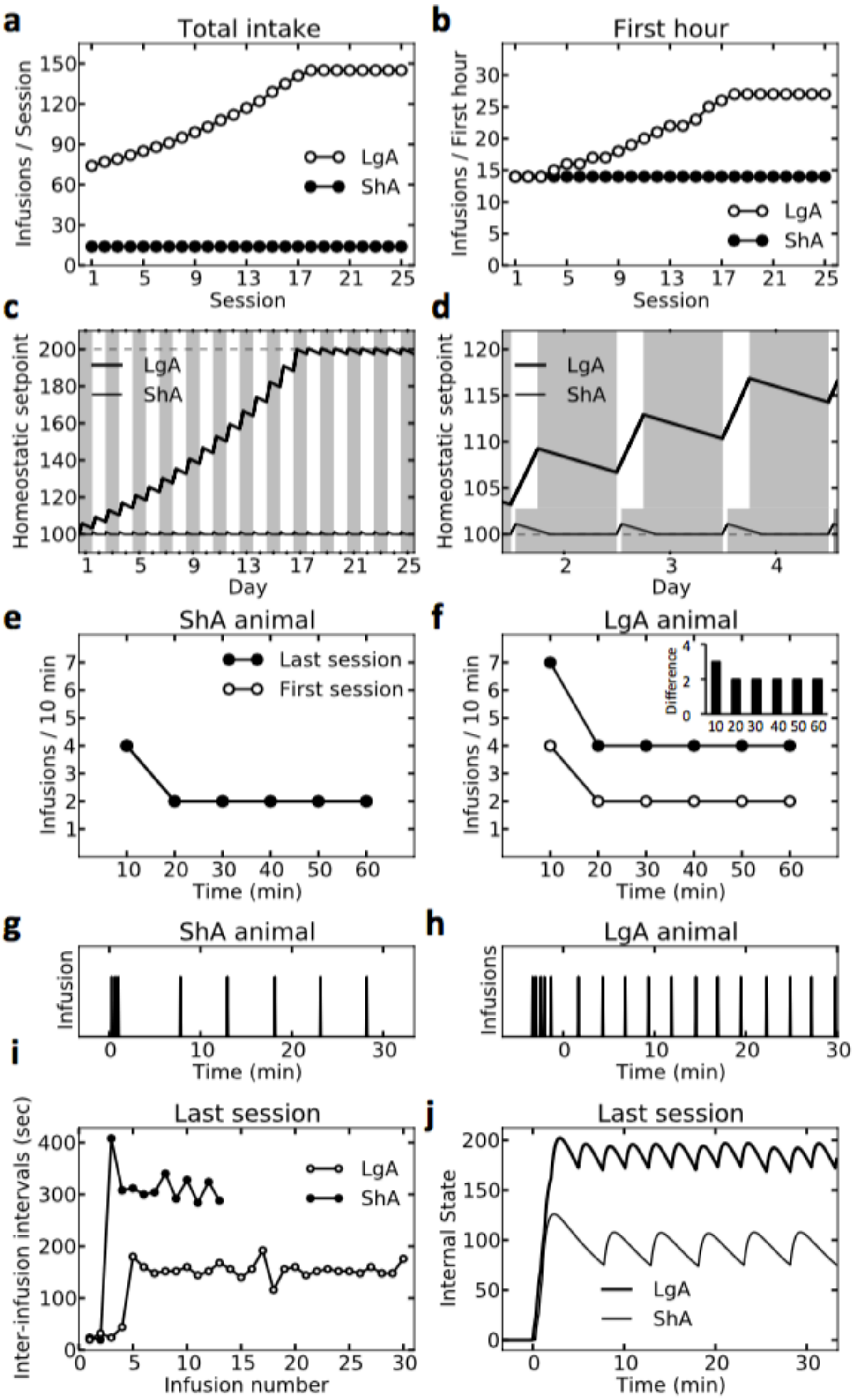

Fig. 2: Simulation results replicating experimental data showing escalation of cocaine self-administration under LgA condition (Fig. S2 and Fig. S3) (S H Ahmed & Koob, 1999; Serge H Ahmed & Koob, 1998). Starting from equal infusion rates in both groups, the LgA agent takes progressively more infusions than the ShA agent (a, b), the underlying mechanism is the gradual rise of the homeostatic setpoint (c). Focusing on the first four trials (d) shows that the the setpoint rise after the 6-hour drug-access period does not recover during the rest of the day (shaded area; 18hrs). Comparing the first and last sessions, the escalation pattern is also

observable in the increased real-time infusion level (over 10min bins) in LgA (f), but not ShA (e) agents. In the LgA agent, this increase is stronger in the first 10min bin of the session as compared to later bins (the inset in plot f; last minus first session). Interestingly we see a loading-effect since the infusion rate in the first 10min block is greater than its steady level, in both ShA (e, g) and LgA (f, h) agents. This "loading effect" comes about because the agents start the sessions in a cocaine-depleted internal state. Thus, reaching the setpoint for the first time (j) requires several infusions with the least possible inter-infusion interval; i.e. 20 seconds (i).

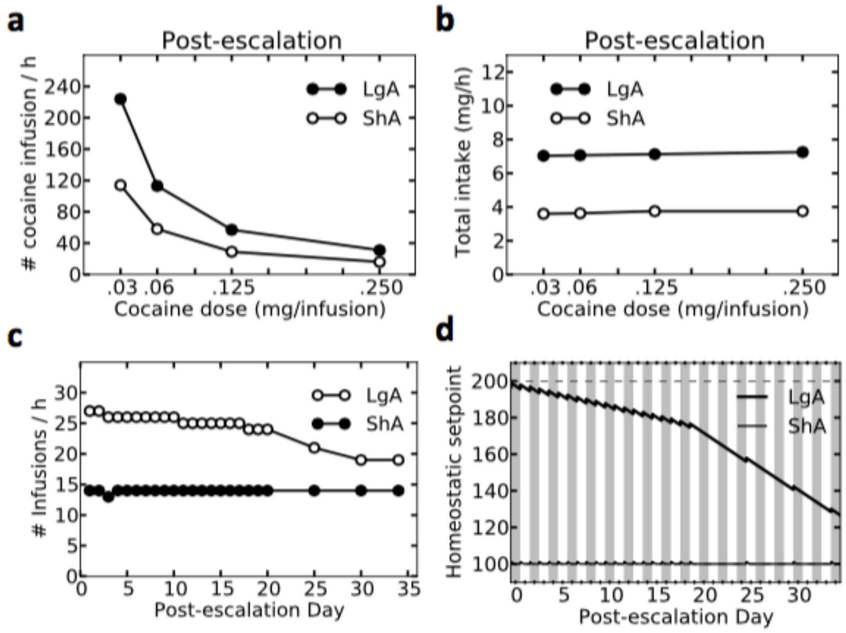

Fig. 3: Simulation results replicating experimental data showing post-escalation dose-dependent cocaine infusion rate decrease and invariant total intake (Fig. S4) (S H Ahmed & Koob, 1999; Serge H Ahmed & Koob, 1998), as well as recovery in LgA animals with post-escalation reduced availability of cocaine (Fig S5) (S H Ahmed & Koob, 1999). (a, b) Post-escalation infusion rate and the total amount of consumed cocaine per hour are higher in LgA than in ShA agents, for all unit doses of cocaine. However, whereas the infusion rate decreases as a function of dose (a), the amount consumed does not change with dose (b). (e, d) After escalation, both LgA and ShA agents are given limited (1hr/day) access to cocaine self-administration. This results in gradual recovery of the setpoint in the LgA agent (c) and thus, in decreasing the rate of infusion (a). After day 20, as in the experiment (S H Ahmed & Koob, 1999), the agents are given only 1hr access to cocaine in every five days. This speeds up the recovery process of the setpoint (b) and thus, accelerates the decreasing trend of infusion rate (a).

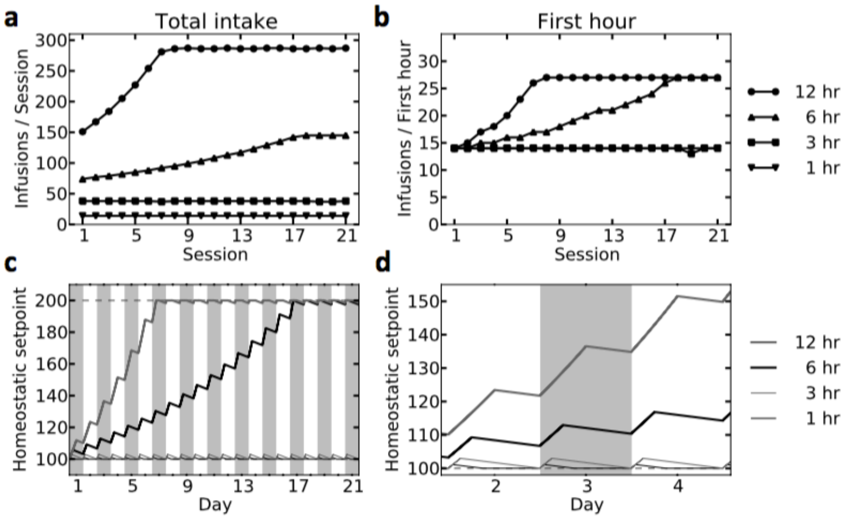

Fig. 4: Simulation results replicating experimental data showing a discontinuous the effect of session-duration on escalation (Fig. S6) (Wee et al., 2007). 1hr and 3hr daily access to cocaine self-administration do not induce escalation (a, b) since even under 3hr access, the elevation of the setpoint is cancelled out during the rest of the day (c, d). Rate of cocaine self-administration increased under 6hr and 12hr access conditions, and this increase was faster in the latter, than in the former (a, b).

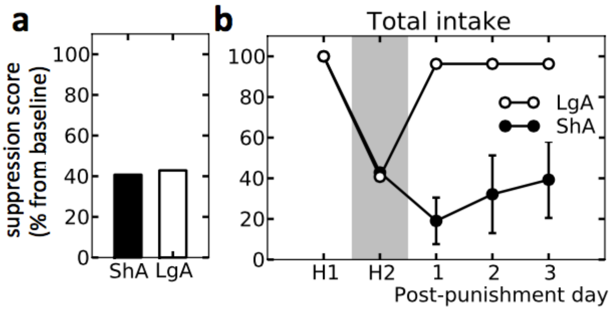

Fig. 5: Simulation results replicating experimental data showing resistance in extended drug-access animals to the punishment-induced suppression of cocaine seeking (Fig. S8) (S H Ahmed, 2012). After 25 days of 6hr vs. 1hr access to cocaine (Fig. 2), both LgA and ShA agents are provided with five sessions of 45min access to cocaine. Only in the second session (indicated by H2 in panel b) cocaine is paired with a punishment. This punishment results in equal immediate cocaine self-administration suppression (from baseline, indicated by H1) in both LgA and ShA agents (a). Whereas the LgA agent rapidly resumes self-administration after removal of the punishment, the ShA agent refrains during at least three consecutive days (b).

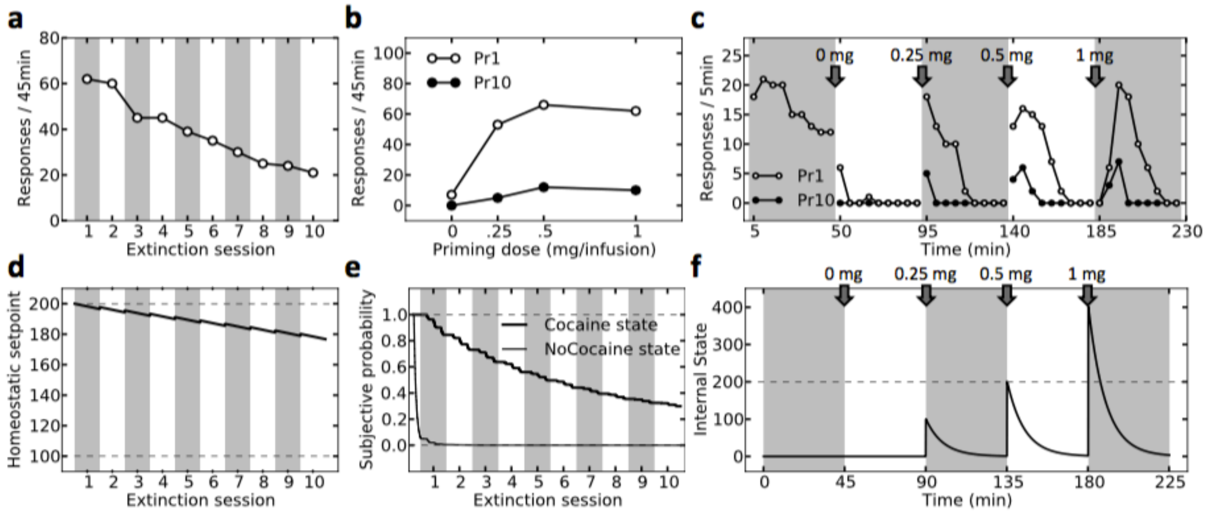

Fig. 6: Simulation results replicating experimental data showing extinction of priming-induced reinstatement (Fig. S9) (Mihindou et al., 2011). After 25 days of 6hr access to cocaine (Fig. 2), the LgA agent undergoes a priming-induced reinstatement procedure during 10 consecutive days. Each day consists of 5 consecutive sessions of 45min, during which pressing the lever has no consequence (extinction). At the beginning of each session, the agent receives a single priming injection of cocaine with the following doses: 0, 0, 0.25, 0.5, and 1mg. (a) The rate of lever-press in the 45min upon infusion of the highest dose (1mg) decreases progressively. (b) Such extinction is seen for various doses of cocaine (compare first (Pr1, white circles) and last (Pr10, filled circles) sessions). (c) Response rates at 5min intervals show a more precise pattern of priming induced reinstatement: transient recovery of infusions at the session start extinguishes progressively during the 10 days. (d) Setpoint decreases but does not recover fully towards control levels in the course of the 10 day extinction procedure. 200 is the escalated level, and 100 is the initial normal level of the setpoint. (e) Extinction of responding over 10 days is due to the decreased subjective probability of receiving cocaine, either when the agent is under cocaine, or when it is not. (f) The extent to which the agent is under cocaine at any time-point replicates experimental data (Frantz et al., 2007) Fig. S9d, e). The dashed line indicates the setpoint level on the first day of the reinstatement experiment.

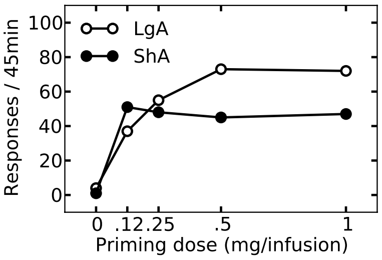

Fig. 7: Simulation results replicating experimental data showing the effect of extended drug-access on dose-dependent priming-induced reinstatement (Fig. S10) (Serge H Ahmed & Cador, 2006). After 25 days of 6hr vs. 1hr access to cocaine, both LgA and ShA agents were provided with one 45-minute session in which, pressing the lever had no consequence. This extinction session was followed by five additional 45-minute extinction sessions, at the beginning of each of which, the agents received different priming doses of cocaine (0, 0.125, 0.25, 0.5, and 1mg). Priming-induced reinstatement was more pronounced in the LgA agent.

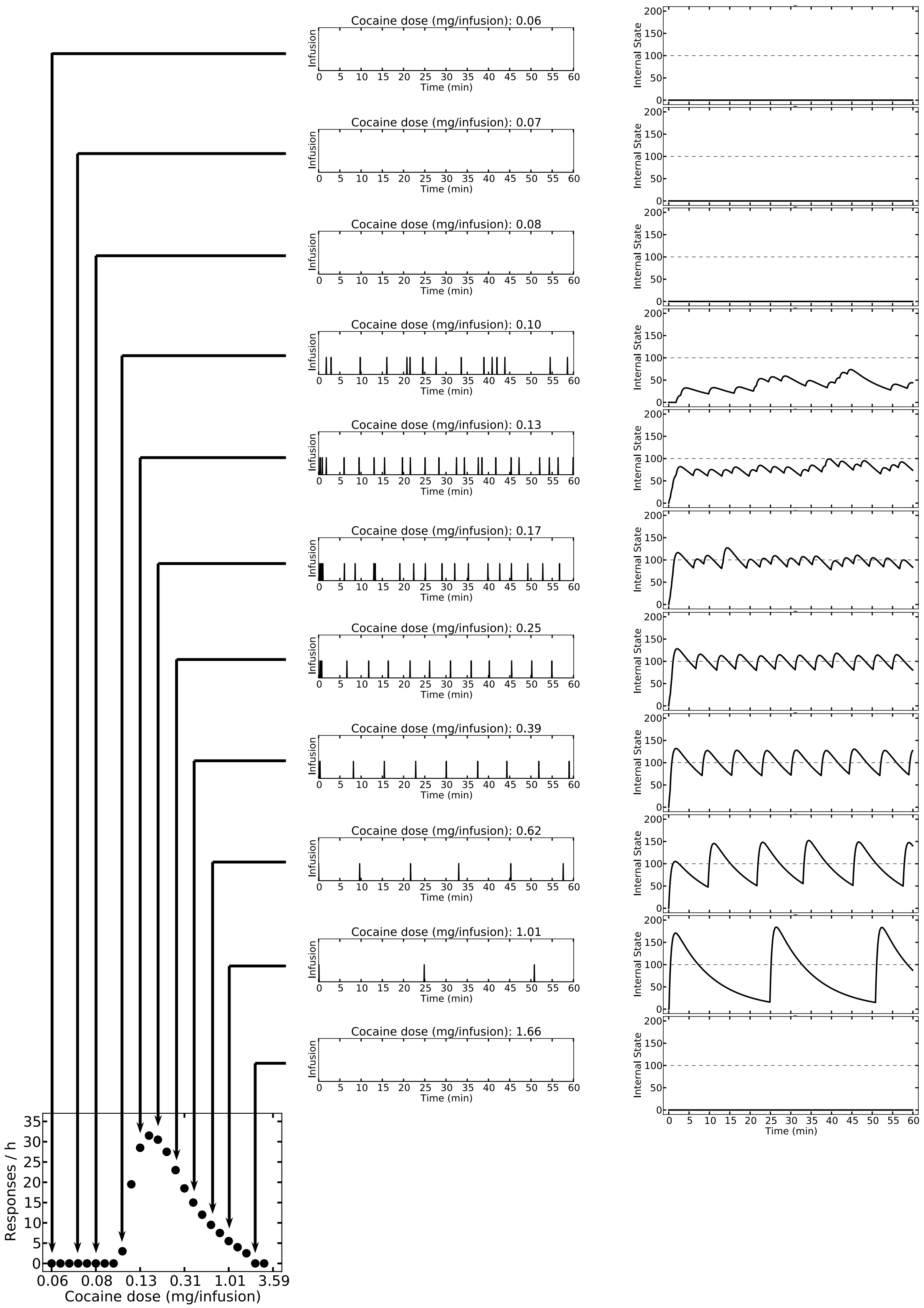

Fig. 8: Simulation results showing the underlying mechanism behind the observed inverted-U dose-response curves. At low doses, the rate of self-administration is low since the cost of

lever-press outweighs the rewarding value of cocaine. As the dose increases, the rewarding value increases and so does the rate of lever-press. At a critical unit dose the rewarding/motivational value is sufficient to evoke a rate of lever-pressing necessary to reach the setpoint. Increasing the dose beyond this critical level, decreases the rate of responding in order to keep the internal state as close as possible to the setpoint (and not overshoot). Sufficiently high doses (the bottom row) result in strong deviations from the setpoint (possibly life-threatening) that the agent learn not to take cocaine.

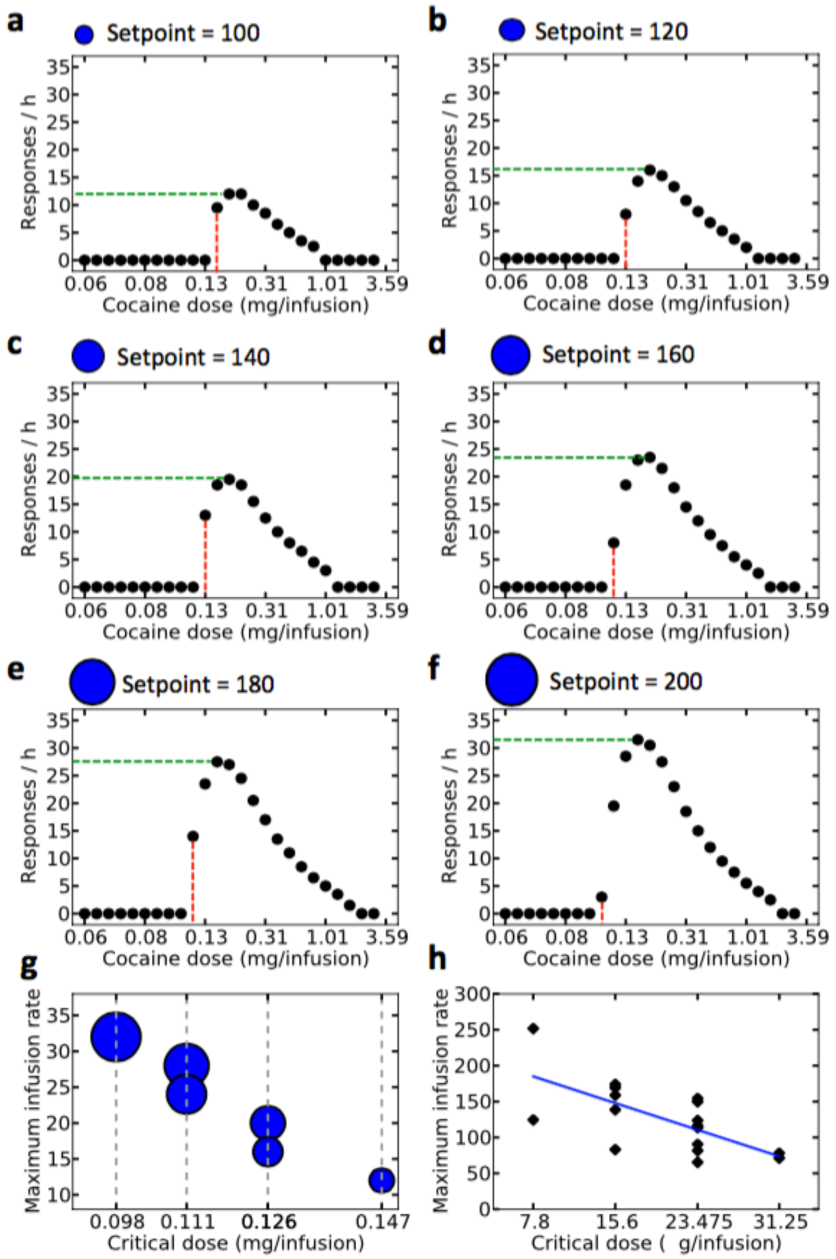

Fig. 9: Simulation and experimental results showing the interaction between critical unit dose and maximum infusion rate. Simulations show that as the setpoint level escalates (increasing order in panels a to f), the minimum unit dose (red line) at which the model shows motivation for seeking cocaine decreases, whereas the maximum infusion rate (green line) among all tested doses increases (simulation results summarized in panel g). (h) Experimental results of cocaine self-administration in rats (n=17) verified the negative correlation between these quantities ($p < 0.01$).

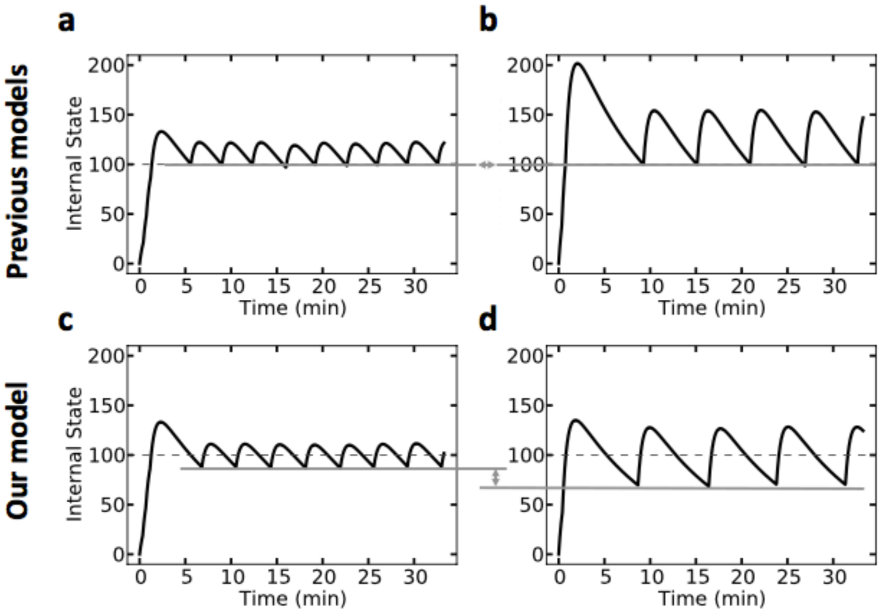

Fig. 10: Simulation results predicting dependence of satiety thresholds on cocaine dose. (a, b) (a,b) Simulations of satiety thresholds for a previously proposed pharmacological model of cocaine self-administration (Serge H Ahmed & Koob, 2005; Tsibulsky & Norman, 1999). In these cocaine-taking response is generated by a slow control signal as the internal state drops below a certain threshold. These models predict that the lower bound of the cocaine level in the brain is equal, for all cocaine doses (compare plots a and b). (c, d) In our model, however, the agent's learned objective is to maintain the internal state as close as possible to the homeostatic setpoint (dashed line in plots c and d). Thus, our model predicts that the lower bound of cocaine level for a low dose of cocaine (c) will be higher than that of a high dose of self-administered cocaine (d).

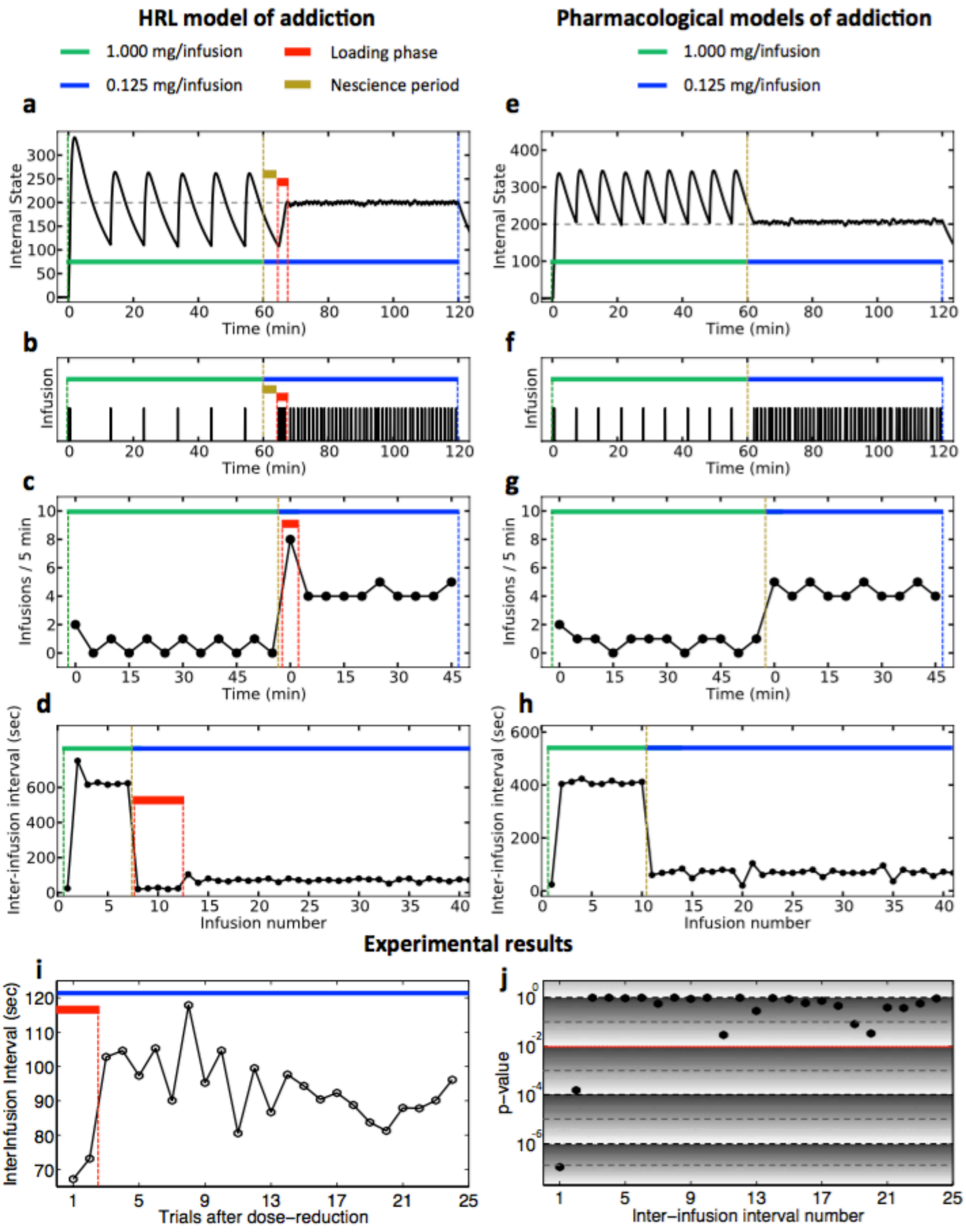

Fig. 11: Effect of within-session reduction of unit-dose on self-administration pattern. After pre-training, both the simulated agents and the rats were tested in a two-hour session where each lever-press resulted in receiving a high vs. low dose of cocaine during the first vs. the second hour of the session. (a-d) Our model predicts a transitory burst of infusion rate after the dose reduction. When the dose is reduced, the agent still waits for a period (nescience) equal to previous inter-infusion intervals so as the internal state drops sufficiently below the setpoint. This is because the agent's objective is oscillate around the setpoint in order to minimize average deviation. Upon the first post-reduction response, the agent detects the increased deviation and, responds intensively in order to catch the setpoint by shortened inter-infusion intervals. After the setpoint is reached, the agent responds with a steady rate just to

oscillate around the setpoint. (e-h) Previous models, in contrast, predict no response burst after dose reduction. In those models, a SA response is elicited every time the internal state drops below the setpoint. After dose-reduction, therefore, as soon as the internal state hits the setpoint, the agent starts responding with a new steady level. Confirming the prediction of our model, experimental results from rats (n=21) showed two significantly shorter inter-infusion intervals (III) right after the dose reduction, as compared to the later IIIs. (i) Average post-reduction III over all rats, over the latest three sessions. (j) *p*-values of one-sided t-tests with the alternative hypothesis that the *i*-th post-reduction III is less than the IIIs between the 10th and the 20th post-reduction infusions (when the response rate is supposedly converged to its new steady level).

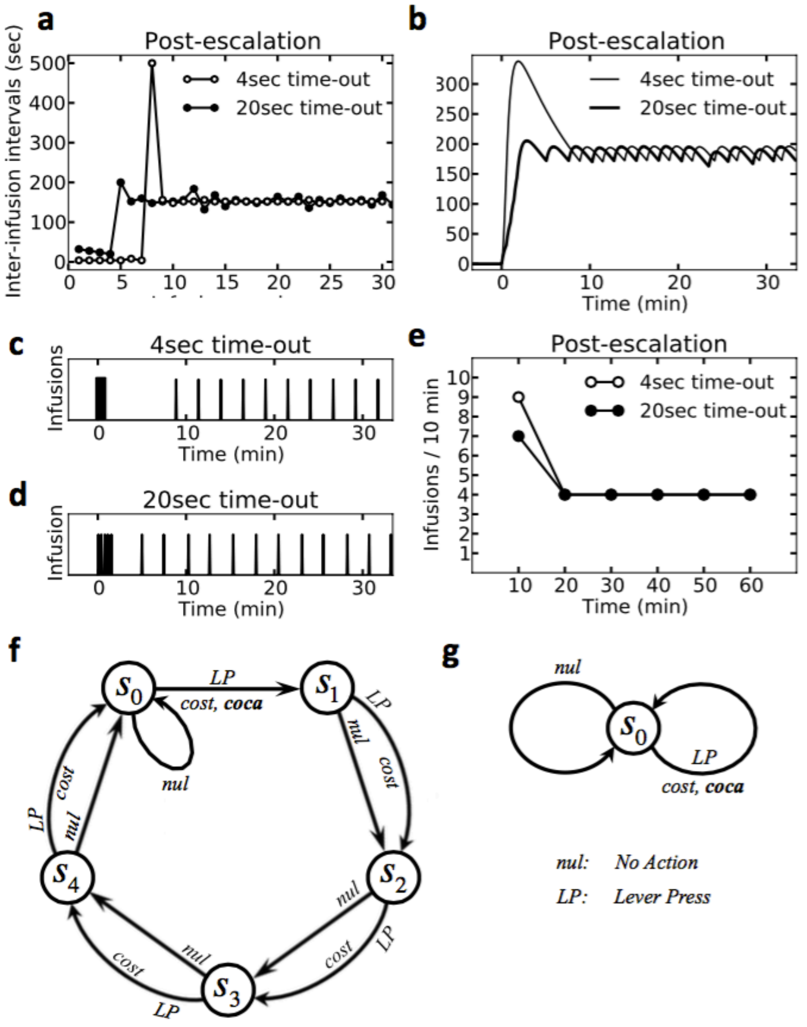

Fig. 12: Effect of time-out duration on the load and pause pattern in self-administration. Simulation results predict more pronounced load and pause effects in the 4sec, compared to the 20sec time-out task. In both 4sec and 20sec cases, the agents start the session in a cocaine-deprived state. Thus, they self-administer cocaine several times, with the lowest possible inter-infusion intervals (a). This period is known as loading phase. However, due to the pharmacodynamics of cocaine, after each infusion, it takes several seconds before cocaine reaches the maximal effect on the internal state (Fig. 1b). For the 20sec time-out, the effect of every cocaine infusion on the internal state is almost completely "sensed" before the next self-administration becomes available. In this condition, the agent's internal state reaches the setpoint after five infusions (b, d). After this loading phase, the agent self-administers steadily. In the 4sec case, however, even though the first few infusions are sufficient to reach the setpoint, their full effect is sensed much later than when self-administration is available again. Thus, the agent continues to take cocaine (a, c) for several extra infusions. These extra infusions result in overshooting the setpoint (b) after their effect on the internal state arrives. In order to return to the setpoint, the agent pauses taking cocaine for several minutes (c), resulting in one significantly large inter-infusion interval (a), known as the pause effect. After that, the agent self-administers steadily. Therefore, the model predicts that both loading and

pause phenomena will more pronounced by decreasing the time-out duration. As in our model the circulating cocaine degrades faster when it is at higher levels, the overshoot of cocaine level in the 4sec case results in more cocaine elimination. In order to compensate for that, the agent takes more infusions of cocaine. As a result, the rate of infusion in the first ten minutes is higher for the 4sec case, than for the 20sec case (e). Plots f and g show the Markov Decision Process used for simulating the 20sec and 4sec cases, respectively.

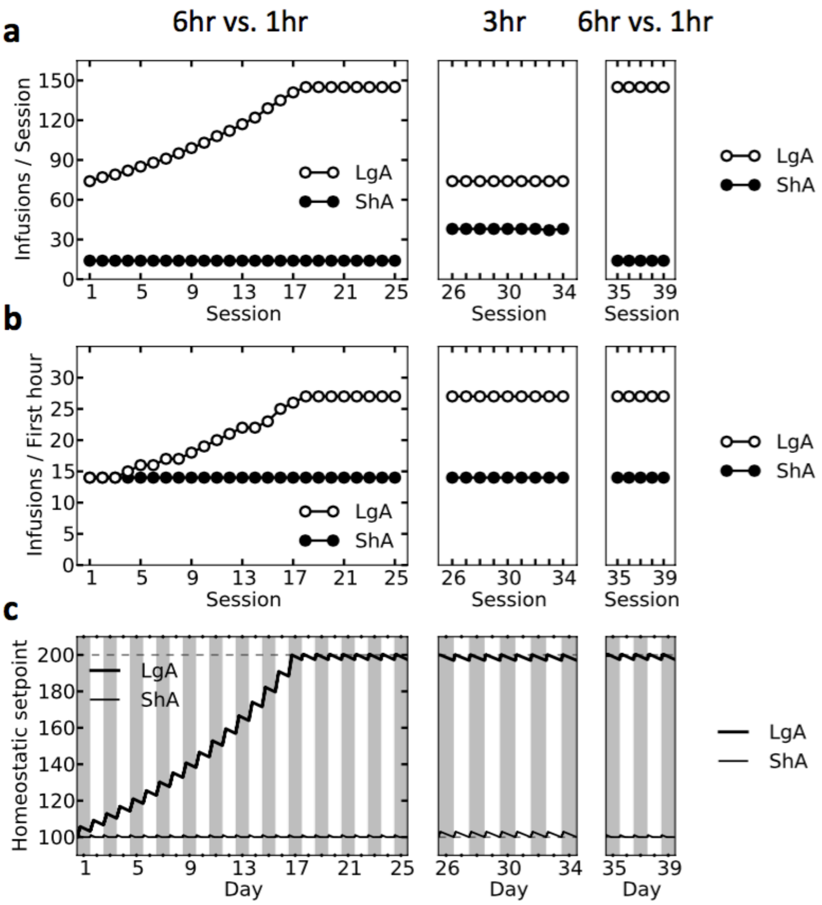

Fig. 13: Simulation results predicting that 3hr access does not induce escalation, but keeps the escalated animals at an escalated level. After 25 sessions of 6hr vs. 1hr access (left panels), both LgA and ShA agents were given 3hr access per day, for 9 consecutive days. The elevation of the setpoint during 3hr is virtually equal to its recovery during the rest of the day (21hr). As these two processes cancel out each other, the setpoint level remains steady under 3hr access condition (plot c, middle). As a result, the rate of infusion/hr remains constant for both ShA and LgA agents (plots a and b, middle). Thus, if after the 3hr access phase, the agents return beck to the 6hr vs. 1hr access conditions, their infusion rate will be equal to the initial steady-state level (panels a and b, right).

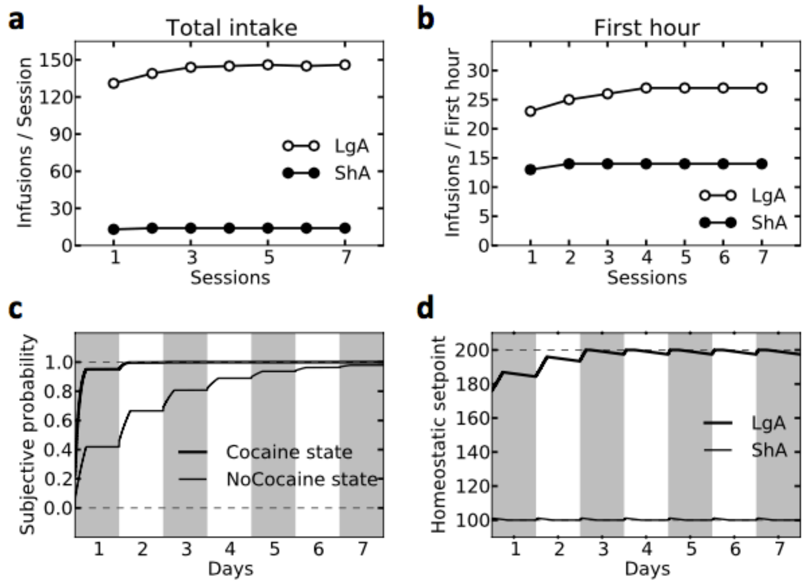

Fig. 14: Simulation results predicting rapid re-escalation of cocaine SA, after extinction of priming-induced relapse. LgA and ShA agent underwent 25 sessions of 6hr vs. 1hr of cocaine self-administration, respectively. They then experienced 10 days in the relapse-extinction schedule as in Fig. 6. After this phase, the agents were again given 7 days of 6hr vs. 1hr access to cocaine self-administration. (a, b) The rate of infusion by the LgA agent remains at an escalated level, even in the first session of re-escalation. This is because the setpoint level was still at an elevated level (d). In fact, the 10-day extinction phase did not lead to recovery of the setpoint (see Fig. 6d), and the extinction of relapse was only due to decreased subjective probability of receiving cocaine (see Fig. 6e). As the subjective probability can be re-learned rapidly within the first session of re-escalation (c), and as the setpoint is still at a high level (d), cocaine infusion rate re-escalates rapidly after extinction of drug-induced reinstatement.

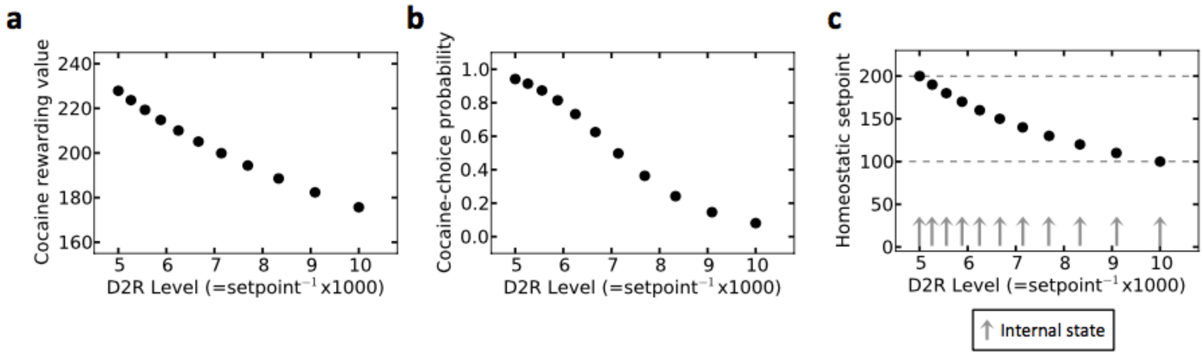

Fig. 15: Simulation results replicating experimental data showing an inverse relationship between drug motivation and D2R availability (Fig. S12) (Nora D Volkow et al., 1999). (a) The rewarding value of a certain unit dose of cocaine decreases as the level of D2 receptor availability increases. (b) ,Probability of the drug-choice outcome is inversely correlated with D2R level, when the choice between is drug or food outcomes. (c) Homeostatic setpoint level is assumed to be encoded inversely by D2R availability ($D2R\ level = 1000.setpoint^{-1}$). The highest and lowest levels of the setpoint are 200 and 100, respectively (equivalent to the D2R levels of 5 and 10, respectively). Choosing the drug option increases the level of the internal state (red arrows) and thus, decreases homeostatic deviation. This drive-reduction reward is higher when the initial distance from the setpoint is higher (i.e., in agent with a high setpoint level, or a low D2R level).

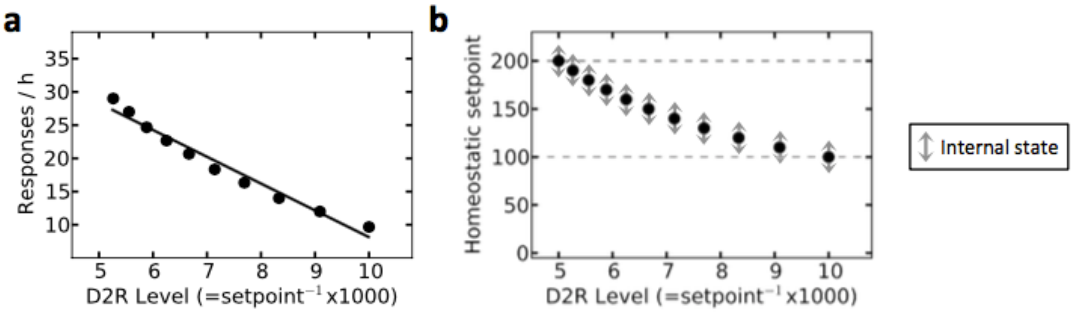

Fig. 16: Simulation results replicating experimental data showing an inverse relationship between cocaine self-administration and D2R levels (Fig. S13) (Michael A Nader et al., 2006). (a) The rate of cocaine self-administration is inversely correlated with D2R level (a). (b) Homeostatic setpoint level is assumed to be encoded inversely by D2 receptor availability ($D2R\ level = 1000. setpoint^{-1}$). The highest and lowest levels of the setpoint are 200 and 100, respectively (equivalent to the D2R levels of 5 and 10, respectively). For each level of D2R availability, cocaine self-administration results in the internal state fluctuating around the homeostatic setpoint (arrows).