LETTER Communicated by A. David Redish

Individual Differences in Nucleus Accumbens Dopamine Receptors Predict Development of Addiction-Like Behavior: A Computational Approach

Payam Piray
piray@ut.ac.ir
Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

Mohammad Mahdi Keramati
keramati_mm@alum.sharif.edu
School of Management and Economics, Sharif University of Technology, Tehran, Iran

Amir Dezfouli
a.dezfouli@ut.ac.ir
Caro Lucas
lucas@ut.ac.ir
Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

Azarakhsh Mokri
mokriazr@sina.tums.ac.ir
Department of Psychiatry, Roozbeh Hospital, Tehran University of Medical Sciences, Tehran, Iran, and Department of Clinical Sciences, Iranian National Center for Addiction Studies, Tehran University of Medical Sciences, Tehran, Iran

Clinical and experimental observations show individual differences in the development of addiction. Increasing evidence supports the hypothesis that dopamine receptor availability in the nucleus accumbens (NAc) predisposes drug reinforcement. Here, modeling striatal-midbrain dopaminergic circuit, we propose a reinforcement learning model for addiction based on the actor-critic model of striatum. Modeling dopamine receptors in the NAc as modulators of learning rate for appetitive—but not aversive—stimuli in the critic—but not the actor—we define vulnerability to addiction as a relatively lower learning rate for the appetitive stimuli, compared to aversive stimuli, in the critic. We hypothesize that an imbalance in this learning parameter used by appetitive and aversive learning systems can result in addiction. We elucidate that the interaction

Mohammad Mahdi Keramati and Amir Dezfouli contributed equally to this work.

between the degree of individual vulnerability and the duration of exposure to drug has two progressive consequences: deterioration of the imbalance and establishment of an abnormal habitual response in the actor. Using computational language, the proposed model describes how development of compulsive behavior can be a function of both degree of drug exposure and individual vulnerability. Moreover, the model describes how involvement of the dorsal striatum in addiction can be augmented progressively. The model also interprets other forms of addiction, such as obesity and pathological gambling, in a common mechanism with drug addiction. Finally, the model provides an answer for the question of why behavioral addictions are triggered in Parkinson's disease patients by D2 dopamine agonist treatments.

1 Introduction

Addiction is a state of compulsive drug-seeking and drug-taking behavior maintained despite its adverse consequences (American Psychiatric Association, 2000). After long-term drug exposure, due to the action of drugs on the brain, individuals become insensitive to the social, behavioral, and health consequences of drug abuse. Nevertheless, this behavior appears only in approximately 20% of individuals initially exposed to the drug (Anthony, Warner, & Kessler, 1994). In humans, clinical observations have shown individual differences in the development of addiction (O'Brien, Ehrman, & Ternes, 1986; de Wit, Uhlenhuth, & Johanson, 1986). Although compulsive behavior appears in animals, all animals that used drugs do not exhibit compulsive behavior (Belin, Mar, Dalley, Robbins, & Everitt, 2008; Deroche-Gamonet, Belin, & Piazza, 2004). Neural and behavioral individual differences underlying vulnerability to addiction are of great importance, and investigation may help identify humans at risk. From a behavioral viewpoint, studies of human addicts and animal models of addiction have implicated some traits, such as impulsivity and novelty seeking, that predate the onset of addiction-like behavior. In a now-classic study Piazza, Deminière, Le Moal, and Simon (1989) reported that novelty seeking in rats predicts development of amphetamine self-administration. Also, impulsive behavior correlates with a tendency to addiction (Koob & Le Moal, 2005). Belin and colleagues (2008) have shown that while novelty seeking predicts a tendency to initiate drug use, impulsivity predicts the development of addiction. Although neural factors that predispose propensity to addiction-like behavior are not fully understood, a wealth of evidence has shown that reduced availability of dopamine receptors in the striatum (Nader, Czoty, Gould, & Riddick, 2008; Volkow, Wang, Fowler, & Telang, 2008), in particular NAc (Dalley et al., 2007; Martinez et al., 2009), correlates with the propensity to increase drug use. More important, using an animal model of compulsive behavior, Everitt and Robbins (2005) and colleagues

have shown that a low availability of NAc D2 receptors predisposes to the development of compulsive behavior (Belin et al., 2008; Dalley et al., 2007).

Experimental studies have shown that another major factor that contributes to the development of addiction is the degree of drug exposure (Deroche-Gamonet et al., 2004; Vanderschuren & Everitt, 2004). In fact, addiction results from the interaction of two factors: the degree of vulnerability and the degree of exposure to drug. There is now wide agreement that the interaction of these two factors impairs the brain's reward learning and memory systems (Everitt et al., 2008; Hyman, Malenka, & Nestler, 2006; Kalivas, 2009; Koob & Le Moal, 2005). Specifically, addictive drugs affect behavior, at least partly, because of their ability to increase synaptic dopamine in the NAc, which plays a critical role in processing natural rewards such as food, as well as in other brain structures that receive dopamine afferents from the ventral tegmental area (VTA), such as the prefrontal cortex (PFC) and amygdala (Robinson & Berridge, 1993). By viewing addiction as a malfunction of the reward learning system and decision making, it should be possible to model this behavioral phenomenon based on decision theory.

Reinforcement learning (RL) theory is perhaps the most popular theoretical framework for decision making in environments that require learning from rewards and punishments. RL has another crucial benefit for modeling addiction: overwhelming evidence supports the hypothesis that dopamine neurons, a major site of action of both natural rewards and addictive drugs, encode prediction error (the difference between the expected and the acquired value of an action), which is a major component of RL theory (Bayer & Glimcher, 2005; Schultz, Dayan, & Montague, 1997). Taking into account these excellent features of RL for modeling addiction, we and others have proposed neurocomputational models to explain possible ways that addiction emerges (Dayan, 2009; Dezfouli et al., 2009; Gutkin, Dehaene, & Changeux, 2006; Redish, 2004; Redish, Jensen, Johnson, & Kurth-Nelson, 2007; Redish & Johnson, 2007; Takahashi, Schoenbaum, & Niv, 2008; Zhang, Berridge, Tindell, Smith, & Aldridge, 2009).

To model the effects of addiction on the reward system requires taking into account different interactions of drug and the dopaminergic circuitry. We can identify four interactions. First, addictive drugs and natural rewards increase the phasic activity of the dopamine signal (Hyman et al., 2006). Although this increased activity can be accommodated in the case of natural rewards, a neuropharmacological noncompensable component in phasic dopamine produced by some drugs, such as cocaine, can lead to addiction (Redish, 2004). Although modeling this drug action explains a possible way that addiction develops, it is not the only reason that these drugs are addictive (Redish, 2004; Redish, Jensen, & Johnson, 2008). Specifically, this cannot be the way that some natural rewards, such as fatty foods and gambling, lead to addiction (Ahmed, 2004). Indeed, overwhelming evidence supports the idea that addiction to both natural rewards and addictive drugs has some common substrates, and this includes similar effects on

the dopaminergic circuit (Potenza, 2008; Volkow et al., 2008). However, natural rewards, unlike drugs, produce a dopamine signal that can be accommodated (Ahmed, 2004). Second, chronic drug consumption causes long-lasting dysregulation of the brain reward threshold (Ahmed & Koob, 1998), which is mainly dependent on the dopamine circuitry (Haber & Knutson, 2009). This persistent dysregulation causes long-lasting changes in processing natural rewards such as sexually evocative visual stimuli (Garavan et al., 2000) and secondary rewards such as money (Goldstein et al., 2007). This action of drugs on dopaminergic circuitry can be related to either regulation of tonic dopamine in NAc or a change in the availability of dopamine receptors in the striatum (Ahmed & Koob, 2005). Third, dopamine receptors within the striatum, especially NAc, play an important role in vulnerability to addiction (Everitt et al., 2008; Volkow et al., 2008). Furthermore, D2 dopamine receptors in the striatum play a key role in the development of addiction to natural rewards (Volkow et al., 2008). Hence, the involvement of the striatal dopamine receptors in vulnerability to the development of addiction is a common component of both drug and food addiction (although the magnitude of their effect can be different). Fourth, although the initial influences of drug are mostly restricted to regions that receive afferents from VTA, including NAc, PFC, and amygdala, chronic exposure affects synaptic dopamine in more dorsal domains of the striatum, which receives dopamine afferents from subtantia nigra pars compacta (SNc), a structure contiguous within the midbrain with the VTA (Everitt & Robbins, 2005; Hyman et al., 2006). In particular, through the development of addiction in susceptible rats, as extensively used in animal models of addiction, the dorsal striatum plays a prominent role in consolidation of drug-seeking habits (Everitt et al., 2008). An important role for the dorsal striatum in human addicts is also reported (Volkow et al., 2006; Wong et al., 2006); however, the ventral- to dorsal-striatum pathway is probably not the sole vulnerable pathway that causes the emergence of addiction (Redish et al., 2008), importantly in humans (Ahmed, 2008). Previously RL models have been proposed by focusing on some of these interactions, such as the effects of drug on phasic dopamine (Dayan, 2009; Dezfouli et al., 2009; Redish, 2004), long-lasting dysregulation of the reward system by drug abuse (Dezfouli et al., 2009; Gutkin et al., 2006), and the distinctive functions of striatum subdivisions and their dopamine sources in development of addiction (Dayan, 2009; Takahashi et al., 2008).

Redish proposed the first RL model for addiction (Redish, 2004) by focusing on the first effect of drugs on dopamine circuitry: neuropharmacological increase in the phasic activity of dopamine. He proposed that a noncompensable component in the prediction error, phasic activity of dopamine signal, results in overvaluation of drug choice and leads to addiction. His model describes a way for the emergence of compulsive behavior: progressive insensitivity of drug choice to the costs associated with drug abuse. We previously extended this model for cocaine addiction by incorporating the

pharmacological effects of cocaine on the threshold of the reward system, using the average reward RL framework (Dezfouli et al., 2009). Taking into account a possible role for tonic dopamine in NAc and also D2 dopamine receptors in the striatum (Ahmed & Koob, 2005), we suggested a possible way that cocaine affects the threshold of the reward processing system. However, the sole reason for cocaine addiction in our previous model is exactly the same as with Redish's model: phasic dopamine remains above zero for a long time and so results in an overvaluation of drug choice. Hence, our previous model extended Redish's approach in modeling cocaine addiction by taking into account the second effect of drugs on the dopamine circuitry. Our approach made it possible to explain some important features of addiction, such as an increased selection of impulsive choice in a delay-discounting task after long-term drug use. However, since the sole reason for cocaine addiction in our previous model is the same as Redish's, our model, like his, cannot account for addictions that are not accompanied by direct pharmacological effects on striatal synaptic dopamine.

Here, we focus on two interactions of drug and dopaminergic circuitry: the role of NAc dopamine receptors in vulnerability to addiction and distinct functions of striatal subdivisions in the development of addiction. Utilizing a modified version of the actor-critic model, a popular RL framework, and modeling the availability of dopamine receptors in striatal subdivisions, we propose a model to explain the role of dopamine receptors and striatal subdivisions in the development of addiction-like behavior. We seek to find an explanation for the fact that development of addiction depends on the interaction of the degree of drug exposure and the degree of individual susceptibility. In addition, we attempt to construct a model that can accommodate both addiction to addictive drugs and natural rewards.

In section 2, we review some neural evidence supporting the core idea of the classic actor-critic model. However, we explain that in contrast to this model, an action-dependent value-learning rule is more consistent with neural observations, and therefore we build our model based on these facts. Then we incorporate the role of NAc dopamine receptors in reward learning into the model, as well as the modifications undergone by these receptors due to the effect of drugs. In section 3, after outlining the theory in a typical compulsivity problem, we explain the major results of the model. Finally, in section 4, we discuss the relation between our model and some previous ones, a behavioral prediction of the model, and some suggestions for future neural experiments and computational models.

2 The Model

2.1 Actor-critic Model. For the purpose of maximizing the accrual of appetitive outcomes and minimizing the aversive ones, the decision maker must learn the consequences of feasible actions by experiencing them in the environment. This acquired knowledge can then direct decisions toward

desirable outcomes in the future. RL theory (Sutton & Barto, 1998) has formulated this concept in the form of a value function that assigns a value to each choice. These assigned values are updated each time the decision maker acquires a certain amount of reward or punishment after performing an action. Action selection in the face of environmental stimuli is based on these estimated values; that is, an action with a relatively high value is more likely to be chosen. Actor-critic is a putative reinforcement learning model that subdivides the decision-making process into prediction and action-selection subtasks, where critic and actor modules are responsible for these two subtasks, respectively (Sutton & Barto, 1998). The two systems are interconnected through a prediction error signal—the difference between the expected and the acquired value of an action. It is becoming increasingly apparent that the striatum and its dopaminergic progressions mediate different functions in terms of the actor-critic model; while the ventral striatum, including NAc, is involved in value learning for the purpose of prediction, the dorsal striatum is responsible for action selection (Montague, Hyman, & Cohen, 2004). These dissociated roles of striatal subdivisions are consistent with a wealth of evidence showing that while the ventral striatum underlies reward processing and motivation (Cardinal, Parkinson, Hall, & Everitt, 2002), the dorsal striatum is involved in motor control (Packard & Knowlton, 2002). Consistently, functional magnetic resonance imaging (fMRI) in humans has measured the blood oxygen level–dependent signal in the NAc during an instrumental conditioning task and its yoked Pavlovian conditioning task. Both tasks require the same learning process, but only the former requires action selection (O'Doherty et al., 2004). This experiment has shown that although the NAc is activated in both tasks (critic), the dorsal striatum is activated only during the instrumental task (actor). Further evidence for striatal functional dissociation in instrumental learning has been found in rats using a reversible lesion technique. Atallah, Lopez-Paniagua, Rudy, & O'Reilly (2007) have shown that while NAc is critical for learning, the dorsal striatum is involved in performance but not learning.

Classically, it is assumed that the actor-critic framework stores state-dependent values in the critic and state-action-dependent preferences in the actor (Sutton & Barto, 1998). The critic underlies learning by computing a prediction error according to its values, and the actor chooses action according to its preferences. Nevertheless, recent instrumental learning studies support the idea that the prediction error underlying learning in instrumental conditioning tasks is computed based on state-action values rather than being action independent. Indeed, most of the earlier studies have used a classical conditioning task or a forced-choice instrumental conditioning task, but not tasks requiring selection among competing actions. In an elegant study, Morris, Nevet, Arkadir, Vaadia, and Bergman (2006) showed that the activity of SNc dopamine neurons in nonhuman primates performing an instrumental conditioning task needed to choose

between two competing choices is modulated by the value of upcoming action. Further evidence supporting the hypothesis that dopamine neurons encode action-dependent prediction error has come from a study by Roesch, Calu, and Schoenbaum (2007), who recorded rats' VTA dopamine neurons during a free-choice instrumental task. In addition to these studies on midbrain dopamine neurons, more recent studies in rats have revealed that some ventral striatal neurons encode information about upcoming action. For example, Kim, Sul, Huh, Lee, and Jung (2009) showed that action values are present in the rats' ventral and dorsal striatal while performing a free-choice task. Stronger evidence has come from a study by Ito and Doya (2009), showed that most information needed for updating a behavioral choice, such as information about reward, action, and state, is present in rats NAc. Most important, Roesch, Singh, Brown, Mullins, and Schoenbaum (2009) aimed at investigating the role of ventral striatum in mediation between cue-evoked activity and decision, have shown that while ventral striatal neurons in a forced-choice task reflect the value of the associated odor cue, which is equivalent to the value of the forced choice, some ventral striatal neurons encode the value of upcoming action in a free-choice task. Indeed, the activity of the ventral striatum is correlated with the integration of value and imminent decision, which is consistent with studies showing that the ventral striatum encodes information about the outcome of choosing an action (Carelli, 2002; Janak, Chen, & Caulder, 2004; Kim et al., 2007; van der Meer & Redish, 2009; Nicola, 2007; Taha & Fields, 2005).

Ito and Doya (2009) and Roesch and colleagues (2009) not only have reported that value representation in ventral striatum is action oriented, but also have confirmed the general role of the ventral striatum in terms of the actor-critic model— its involvement in learning but not action selection. Roesch et al. (2009) reported that the values of unselected actions are not represented in the ventral striatal neurons and thus, the information in this structure is not sufficient for action selection. Unlike this observation, increasing evidence has shown that the values of actions, even those that are not selected, are encoded by dorsal striatal neurons (Lau & Glimcher, 2008; Pasquereau et al., 2007; Samejima, Ueda, Doya, & Kimura, 2005). For example, Samejima et al. (2005) have shown that dorsal striatum encodes action values while monkeys should choose between two different actions. Consistently, Ito and Doya (2009) have reported that information related to action choice is very weak in the ventral striatum before the onset of the choice, but increases rapidly after choice onset. Taking into account their previous study (Samejima et al., 2005) and also other studies (Lau & Glimcher, 2008; Pasquereau et al., 2007), they concluded that although action values are present in the ventral striatum, action-value coding neurons are less dominant in the ventral, as compared to dorsal, striatum, and so the information encoded in the ventral striatum can be useful for action evaluation and learning rather than action selection.

In line with these observations, we assume that the ventral and dorsal striata serve different functions in instrumental conditioning tasks. Moreover, we assume that the critic computes action-dependent prediction error, which in turn has an impact on downstream dopamine neurons that subsequently modify both the critic and the actor. In formal language, supposing that the decision maker leaves its current state, s_t, to the next state, s_{t+1}, after choosing an action, a, and obtaining a reward, r_t, the prediction error signal can be calculated as

$$\delta = r_t + \gamma V(s_{t+1}, b) - V(s_t, a), \tag{2.1}$$

where $V(s_t, a)$ is the critic's expected value for the action a in state s_t, γ is the discounting factor, which indicates the relative incentive value of immediate compared to delayed rewards, and b is the action with maximum value in the next state.

Within this computational framework, the prediction error signal is employed for updating the critic's value, as well as updating the behavioral preference of the decision maker, stored by the actor:

$$w_V(a) \leftarrow w_V(a) + \alpha\delta s_t \tag{2.2}$$

$$w_P(a) \leftarrow w_P(a) + \alpha\delta s_t, \tag{2.3}$$

where $w_V(a)$ and $w_P(a)$ are the weights for a linear estimation of the critic's values and actor's preferences, respectively (McClure, Daw, & Montague, 2003; Montague, Dayan, & Sejnowski, 1996). α is the learning rate, describing the degree to which the prediction error signal affects values and preferences. In general, s_t can be a vector that represents different stimuli. Here we use binary representation, in which the ith element of the s_t is 1 if and only if the stimulus i is present; otherwise this element is equal to 0. Apparently if the simulated environment has only one stimulus, we simply have $s_t = 1$ whenever that stimulus is present and $s_t = 0$ otherwise. The critic's values and actor's preferences are then calculated according to their weights:

$$V(s_t, a) = w_V.s_t \tag{2.4}$$

$$P(a \mid s_t) = w_P.s_t, \tag{2.5}$$

where $P(a \mid s_t)$ is the actor's behavioral preference of choosing a when the decision maker is in state s_t. Although neural evidence supports the idea of the existence of two separate action-dependent structures for learning and action selection, the benefit of the existence of two separate structures for a pair of stat-action is not clear from a normative viewpoint. Roesch et al. (2009) have suggested a broad idea for describing the difference in representation of action in these two structures. They suggested that while

neuronal activity showing actions' preferences in the dorsal striatum is biased according to the neurons' anatomy and the direction of response coded by them, neural activity showing actions' values in ventral striatum is less dependent on a particular spatial response. For example, while neurons in the occulomotor caudate typically show greater activity for saccades made in the direction opposite the recording hemisphere (Lauwereyns, Watanabe, Coe, & Hikosaka, 2002), almost an equal number of neurons in the ventral striatum prefer rightward and leftward movement (Roesch et al., 2009). Further investigations are required to answer this question exactly.

Finally, having preferences for actions in each state, the actor selects a choice using the softmax rule:

$$\pi(a) = \frac{e^{\beta P(a|s_t)}}{\sum_b e^{\beta P(b|s_t)}}, \tag{2.6}$$

where $\pi(a)$ is the probability of choosing a, and β is a factor that determines the rate of exploration.

2.2 The Proposed Model. The striatum and its connections with midbrain dopaminergic neurons play a critical role in drug taking and drug addiction. In order to investigate the role of these brain regions in susceptibility to addiction and compulsive drug seeking, a more elaborate model than the simple actor-critic is needed. In this section, we incorporate more detailed neural findings into the actor-critic model of the striatum to explain a wide variety of evidence from the neuroanatomy of the striatum and its connections with the midbrain dopaminergic neurons. In taking into account the role of the NAc dopamine receptors at different stages of addiction and also addressing vulnerability to addiction, we slightly modify equation 2.2 to capture the function of NAc dopamine receptors.

2.2.1 The NAc Shell and Posteromedial VTA Play a Critical Role in Initial Phases of Addiction. NAc can be divided histologically and anatomically into core and shell regions. It has been suggested that drugs affect responses by increasing synaptic dopamine in the NAc, especially within the shell region (Di Chiara et al., 2004; Hyman, 2005; Ikemoto, 2007). For example, rats easily learn to self-administer cocaine and amphetamine into the shell, but not the core (Ikemoto, 2003; Ikemoto, Qin, & Liu, 2005; Rodd-Henricks, McKinzie, Li, Murphy, & McBride, 2002). Intracranial self-administration with a mixture of D1 and D2 receptor antagonists shows the critical role of the NAc shell in mediating the rewarding effects of drugs (Shin, Qin, Liu, & Ikemoto, 2008). Moreover, infusions of dopamine receptor agonists and antagonists into the shell alter the rate of drug self-administration in the rats (Carlezon, Devine, & Wise, 1995; Fenu, Spina, Rivas, Longoni, & Di Chiara, 2006; Ikemoto, Glazier, Murphy, & McBride, 1997; Spina, Fenu, Longoni,

Rivas, & Di Chiara, 2006). Similar results have also been demonstrated by lesion studies (Sellings, McQuade, & Clarke, 2006b, 2006a; Sellings & Clarke, 2003). Anatomically, the NAc shell receives dopaminergic projection from posteromedial VTA (Ford, Mark, & Williams, 2006; Ikemoto, 2007). This part of the VTA is mostly affected by appetitive, but not aversive, stimuli (Bolanos et al., 2003; Carlezon et al., 2000; Ikemoto, 2007; Olson et al., 2005). This topographical difference within the VTA is important for mediating the rewarding effects of many drugs such as cocaine, nicotine, carbachol, cannabinoids, ethanol, and opiates (Carlezon et al., 1997, 2000; Ford et al., 2006; Ikemoto, Murphy, & McBride, 1998; Ikemoto et al., 1997; Ikemoto & Wise, 2002; Rodd et al., 2004, 2005; Zangen, Ikemoto, Zadina, & Wise, 2002; Zangen, Solinas, Ikemoto, Goldberg, & Wise, 2006) and also for natural rewards such as sucrose (Bolanos et al., 2003).

2.2.2 The NAc Receptors Modulate the Rewarding Effect of the Drug and Predict Vulnerability to Addiction. NAc dopamine receptors play a crucial role in different stages of drug use. For cocaine, drug self-administration is strongly regulated by both D1-like (hereafter D1) and D2-like (hereafter D2) receptors (Self, 2004). Cocaine taking diminishes by pretreatment with either D1 or D2 receptor agonists (Caine, Negus, Mello, & J. Bergman, 1999), while pretreatment with either D1 or D2 receptor antagonists increases cocaine taking (Corrigall & Coen, 1991). Neuroimaging studies have demonstrated a crucial role for D2 receptors, specifically in the NAc, in propensity to drug use. In human subjects, individuals with lower D2 receptor availability report psychostimulants to be pleasant, while individuals with higher D2 receptors find the same psychostimulants aversive (Volkow et al., 1999). Positron emission tomography (PET) studies in nonalcoholic human subjects with a dense family history for alcoholism have shown that they have higher D2 receptors in the striatum than individuals without such family histories (Volkow et al., 2006). In nonhuman primates, cocaine self-administration propensity is negatively correlated with the availability of D2 receptors and also social dominance (Morgan et al., 2002; Nader et al., 2006). Consistently, increased D2 receptors in rat NAc markedly reduce alcohol self-administration (Thanos et al., 2001). Recently and more important, it has been reported that D2 receptors availability in rat NAc, but not dorsal striatum, predicts the escalation of cocaine self-administration and progression to compulsive behavior (Belin et al., 2008; Dalley et al., 2007).

Although evidence for the role of D1 receptors in predisposition to addiction is limited, low D1 receptor availability in the ventral striatum recently has been reported to be associated with the choice to self-administer cocaine (Martinez et al., 2009). Moreover, extensive evidence has shown that coactivation of both D1 and D2 receptors in the NAc, especially the shell region, is necessary for mediating the rewarding effects of drugs at the initial stages of drug use (Bachtell, Whisler, Karanian, & Self, 2005; Edwards, Whisler, Fuller, Orsulak, & Self, 2007; Hopf, Cascini, Gordon, Diamond, &

Bonci, 2003; Inoue et al., 2007). For example, rats self-administer both D1 and D2 receptor agonists into the NAc shell in combination but not alone (Ikemoto et al., 1997). In addition, the effect of dopamine in NAc shell is inhibited by either D1 or D2 antagonist (Inoue et al., 2007).

In sum, dopaminergic projections from posteromedial VTA to NAc shell are involved in appetitive but not aversive learning and regulate the postsynaptic activity of dopamine receptors in the NAc shell. These projections mediate the prediction error signal used for instrumental learning in the downstream circuits. Computationally, the critic (i.e. NAc shell) uses the prediction error signal, carried by the phasic activity of dopamine neurons originating from posteromedial VTA, to update its value estimations. The prediction error is itself modulated by the availability of dopamine receptors in the critic (NAc shell) in the case of appetitive stimuli. Hence, we substitute equation 2.2 with

$$w_V(a) \leftarrow w_V(a) + \kappa_c \alpha \delta s_t \quad \text{if } r > 0, \tag{2.7}$$

where κ_c is the normalized availability of dopamine receptors in the NAc shell, which is equal to 1 in normal situations. We do not dissociate the function of D1 and D2 receptors in equation 2.7 because evidence supports the hypothesis that decreased availability of either D1 (Martinez et al., 2009) or D2 (Dalley et al., 2007) receptors in the NAc predisposes a propensity to developing addiction. More important, cooperative activity of both receptors in the NAc shell is critical for mediating the rewarding effects of the drug (Hopf et al., 2003; Ikemoto et al., 1997; Self, 2004), and so lower availability of either D1 or D2 influences the rewarding effects of the drug. It is noteworthy that the functional role of D1 and D2 receptors in more dorsal striatal organizations is likely different, especially in mediating appetitive and aversive learning (Hazy, Frank, & O'Reilly, 2007). By modeling dopamine receptor availability in the NAc shell, we define vulnerability to addiction by $\kappa_c < 1$, that is, a decreased availability of D1 or D2 receptors within the NAc shell.

Since the posteromedial VTA dopamine neurons are involved in only appetitive but not aversive learning, the prediction error for aversive stimuli is not modulated by κ_c, and the original prediction error signal will be used directly for the purpose of value learning. Though serotonergic system has been proposed as a candidate for handling aversive outcomes of decisions (Daw, Kakade, & Dayan, 2002; Dayan & Huys, 2008), regardless of what neuronal circuit is responsible for handling it, we use equation 2.2 in the case of $r < 0$.

2.2.3 Long-Term Drug Use Affects Availability of Dopamine Receptors in the NAc. Chronic exposure to a drug affects dopamine receptor availability within the striatum. Human subjects with a wide range of drug addictions

have shown significant reductions in D2 receptor density within the striatum, including the NAc and the dorsal striatum (Volkow, Fowler, Wang, & Swanson, 2004). In nonhuman primates, D2 receptor reduction is initially significant in ventral regions, but by chronic exposure to the drug, the effects of cocaine spread dorsally to include the dorsal striatum (Nader et al., 2002; Porrino, Daunais, Smith, & Nader, 2004). Nevertheless, the effect of the drug on D1 receptor density is still a matter of significant debate. Animal studies have shown both an increase (Nader et al., 2002) and a decrease, particularly in the shell region (Edwards et al., 2007; Moore, Vinsant, Nader, Porrino, & Friedman, 1998), in striatal D1 receptor availability after long-term cocaine use.

To model the dynamics of receptor availability in NAc, which plays a critical role in addiction, we assume that the availability of dopamine receptors in NAc decreases with drug use. This assumption is consistent with evidence for D2 receptor changes by drug use and also with the critical role of coactivation of D1 and D2 receptors within the NAc shell for mediating the rewarding effects of drug and natural rewards (Hopf et al., 2003; Ikemoto et al., 1997; Self, 2004). D2 receptors' reduction depends on three factors: the receptors' initial density, dopamine release, and the duration of drug exposure (Laruelle et al., 1997; Mach et al., 1997; Porrino et al., 2004). Hence, we model the reduction of the NAc shell receptors by slightly reducing κ_c by each experience with drug. We propose that κ_c decreases exponentially, depending on dopamine release in the NAc shell, duration of drug exposure, and current density of receptors. In fact, after each action, the amount of dopamine release in the critic, which disruptively affects dopamine receptors, is computed as

$$\delta_{cd} = \begin{cases} \delta - \kappa_c \delta_0 & \text{if } \delta_c > \kappa_c \delta_0 \\ 0 & \text{if } \delta_c < \kappa_c \delta_0 \end{cases}, \tag{2.8}$$

where δ_0 is a constant threshold. Then κ_c will be reduced according to the calculated disruptive amount of dopamine release:

$$\kappa_c \leftarrow \kappa_c e^{-\delta_{cd}/\tau}, \tag{2.9}$$

where τ is a constant that determines the degree of influence of δ_{cd} on κ_c. δ_0 in equation 2.8 is a free parameter of the model, which determines the range of dopamine release that is able to decrease receptor density. Obviously this free parameter should be more than the maximum possible natural reward; otherwise, the natural reward will induce reduction in the availability of receptors under normal conditions, that is, $\kappa_c = 1$. Another factor that affects receptor density is the current density of the receptors, and so the initial availability of receptors influences the rate of receptor reduction. Therefore, δ_{cd} will be zero for unsusceptible subjects in the face

of the highest possible natural reward and, hence, no reduction will happen to the availability of receptors. For the normal release of dopamine, the availability of receptor remains unchanged. However, stimuli with abnormally high rewarding effects reduce the level of receptors in the NAc shell. This reduction by itself intensifies the rate of receptor reduction and, hence, subsequent consumption of the reward will decrease the availability of receptors further. Finally, after long-term abstinence, the dopamine receptors recover to their original level (Beveridge, Smith, Nader, & Porrino, 2009). This recovery of dopamine receptors and perhaps its dynamics might be significant to relapse after long-term extinction. However, since we do not focus on this feature of addiction, modeling the recovery of receptors is not required here.

2.2.4 Dorsal Striatum Plays a Prominent Role in Later Phases of Addiction. Although the initial reinforcing effects of drugs depend on the NAc, the consolidation of rigid habitual responses and ultimately compulsive drug-seeking behavior depends on the dorsal striatum and SNc (Berke & Hyman, 2000; Everitt et al., 2008; Everitt & Robbins, 2005; Hyman et al., 2006). For example, the rat dorsal striatum and its dopaminergic innervations play an important role in cocaine seeking before cocaine self-administration (Ito, Dalley, Robbins, & Everitt, 2002). Furthermore, blockade of dopamine in the rat dorsal striatum inhibits cocaine seeking in a second-order schedule of reinforcement (Vanderschuren, Di Ciano, & Everitt, 2005). The progressive involvement of dorsal striatal-dependent processes in cocaine-self administration has also been shown in metabolic and molecular imaging studies of nonhuman primates (Porrino, Lyons, Smith, Daunais, & Nader, 2004). In humans, an increase in dopamine release in the dorsal striatum in cocaine-addicted individuals watching cocaine cues has been reported (Volkow et al., 2006). Hence, it is hypothesized that the behavioral transition from goal directed to habitual and compulsive drug seeking is based on a transition of control over drug-seeking behavior from the ventral striatum and prefrontal structures to the dorsal striatum, which is mainly mediated by striatal dopaminergic innervations (Everitt & Robbins, 2005). This theory is inspired by studies in animal behavioral psychology showing that the dorsal striatum plays a critical role in habitual behavior (Balleine, Delgado, & Hikosaka, 2007). Nevertheless, there is no agreement for the role of D1 and D2 receptors (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007; Noble, 2003) or the involvement of SNc dopamine neurons in appetitive and aversive learning (Bayer & Glimcher, 2005; Matsumoto & Hikosaka, 2009), as well as for the dynamics of D1 receptors' changes by drug use (Edwards et al., 2007; Nader et al., 2002). Hence, we do not incorporate the role of dorsal-striatal receptors into the model to explain all questions noted in section 1. However, there is some evidence for distinct functions of D1 and D2 receptors in the NAc shell, core, and dorsal striatum (Hazy et al., 2007; Pezze, Dalley, & Robbins, 2007). Additionally, although it has been shown

that long-term drug abuse decreases the availability of D2 receptors in the dorsal striatum (Nader et al., 2002), its effect on D1 receptors is controversial (Edwards et al., 2007; Nader et al., 2002). Moreover, it has been proposed that D2 receptors in the dorsal striatum are involved in learning from negative but not positive outcomes (Hazy et al., 2007), which is compatible with recent evidence on Parkinsonism (Frank, Seeberger, & O'Reilly, 2004) and D2 gene polymorphism studies (Frank, Doll, Oas-Terpstra, & Moreno, 2009; Frank et al., 2007; Klein et al., 2007). We show later that incorporating the changes of D2 dorsal-striatal receptors into the model not only maintains the ability of the model in describing the addiction, but also enhances the model to explain some surprising findings around addiction and Parkinsonism (Dagher & Robbins, 2009).

3 Results

3.1 Theory Outline. Here, we explain how our formulation, under certain conditions, can result in nonoptimal behavior, as in addicts. Mathematically, the nonoptimal behavior is related to different learning rates of the critic and the actor in the face of appetitive and aversive stimuli. In our model, the learning rate for aversive learning in both critic and actor is α. It means that the critic's value and the actor's preference are updated each time by an equal amount, $\alpha\delta$, in the face of aversive stimuli. On the other hand, while the learning rate for appetitive stimuli in the critic is $\alpha\kappa_c$, the learning rate for the actor is α, and so the critic's value is updated by $\alpha\kappa_c\delta$, whereas the actor's preference is updated by $\alpha\delta$. These different learning rates for appetitive stimuli in addition to equal learning rates for the aversive stimuli have two important effects. The first is an obvious transitory effect: the speed of learning for appetitive and aversive learning is different. The second effect is a persistent bias that plays a prominent role in emergence of nonoptimal behavior: the actor's preferences are abnormally exaggerated in the face of appetitive but not aversive stimuli.

It is notable that the difference between appetitive and aversive learning rates in a temporal difference RL model (but not actor-critic) does not result in a permanent bias. Also, dissociating the learning rates of actor and critic in the actor-critic model in which appetitive and aversive systems are not dissociated does not result in persistent nonoptimal action selection; it affects the speed of learning that can be resolved by enough exploration. Indeed, both conditions are required for nonoptimal biased behavior to emerge. Mathematically, the nonoptimal behavior appears if

$$\frac{\text{Actor's learning rate for appetitive stimuli } (\alpha)}{\text{Critic's learning rate for appetitive stimuli } (\alpha\kappa_c)} \neq \frac{\text{Actor's learning rate for aversive stimuli } (\alpha)}{\text{Critic's learning rate for aversive stimuli } (\alpha)}.$$

As explained earlier, in the light of our model, a vulnerable individual has a reduced density of dopamine receptors in the NAc—$\kappa_c < 1$. Hence, for a vulnerable individual, the left ratio is larger than the right one, and so the selected actions are maladaptively biased toward appetitive stimuli.

Now assume a simple environment that consists of one state, s, where one lever is available and the drug reward is obtained by the lever-press action, a (see Figure 1a, Phase 1). In each time step, the model chooses between action a and a no-reward action b (doing nothing). We want to investigate the effects of drug on a vulnerable model, $\kappa_c < 1$. By pressing the lever, the rat takes the drug, and the correspondingly high level of dopamine, δ, is released according to equation 2.1. This abnormally higher prediction error, δ, decreases κ_c by chronic drug use according to equations 2.8 and 2.9. This process will be launched only for susceptible individuals with an abnormally lower density of NAc receptors. Furthermore, while the weight of the value, $w_V(a)$, is updated by $\alpha\kappa_c\delta$ (see equation 2.7), the weight of the preference, $w_P(a)$, is updated by $\alpha\delta$ (see equation 2.3). Due to the fact that $\kappa_c < 1$, the preference for pressing the lever, $P(a \mid s)$, grows more than its value, $V(s, a)$. Because of the diminishing trend of κ_c, the preference for taking the drug is progressively augmented by the development of addiction. Since choosing b leads to no reward and no punishment, its value, $V(s, b)$, as well as its preference, $P(b \mid s)$ remain zero.

Now assume that the drug reward is removed, and pressing the lever is associated with an acute shock punishment, $r_{sh} \ll 0$ (see Figure 1a, phase 2). We now expect an intact model to choose the no-reward action, b, which is much better than action a, which leads to the acute punishment. However, this is not the case for a vulnerable model (see Figure 1b). Due to normal learning of aversive stimuli in both critic and actor (see equations 2.2 and 2.3), for a vulnerable model, pressing the lever results in an equal decrease in value and preference of action a, both equal to $\alpha\delta$. After a while, although the value of this state action, $V(s, a)$, becomes negative, but because $P(a \mid s) \gg V(s, a)$ before the advent of punishment, the preference for pressing the lever remains positive, and thus $P(a \mid s) > P(b \mid s) = 0$. Hence, the lever press action is chosen by the actor more than the no-reward action, even in the face of aversive outcome. Through the course of learning, the prediction error for the lever press converges to 0 (because $V(s, a) \to r_{sh}$), and so the actor's preference for this action remains positive. Notably, because the reward associated with b is 0, its value and preference are always 0.

In other words, although it is expected that an acute punishment will be able to cancel out the rewarding consequences of drug taking and the decision maker to stop using drug, this is not how a vulnerable model behaves. Despite the fact that the value of drug seeking is negative, the preference for selecting drug remains positive. This discrepancy between the covert value of drug seeking and the overt behavior of the individual is due to the assumption that aversive and appetitive stimuli have an asymmetric impact on action selection policy. For an aversive stimulus, because both

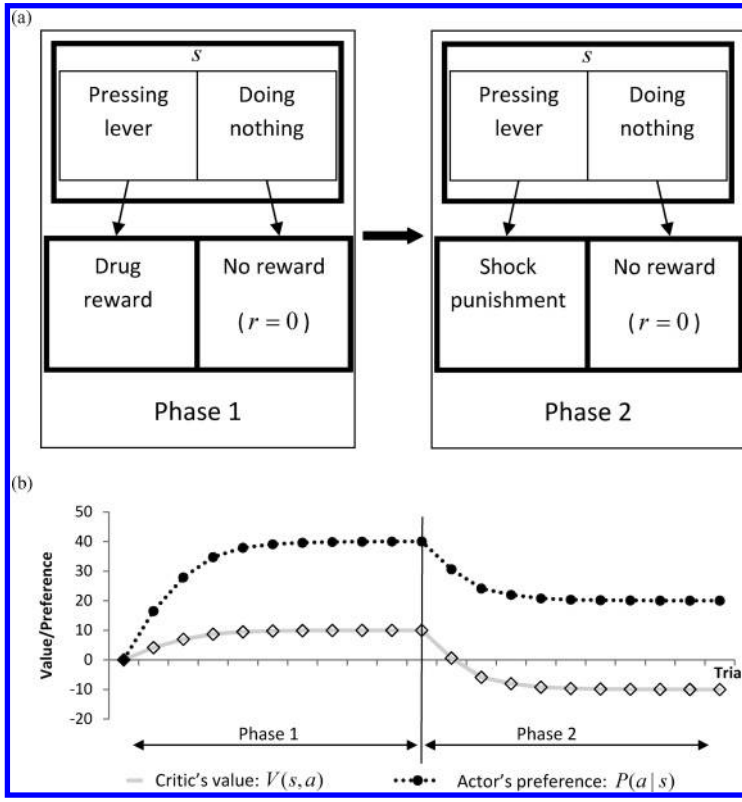

Figure 1: (a) A vulnerable model with $\kappa_c = 0.25$ performs the task depicted in phase 1. In state s, the model chooses between two actions. While action a results in drug reward, $r_d = 10$, action b results in no reward. After training in this phase, the drug reward is substituted with a shock punishment, $r_{sh} = -20$(phase 2). (b) The performance of the vulnerable model in the environment. While the optimal behavior in phase 2 is choosing action b, the vulnerable model chooses action a in the face of punishment (because $P(a \mid s) > P(b \mid s) = 0$). It is because the preference toward action a in phase 1 is exaggerated in the actor, while it is normal in the critic. Since both value and preference are updated by an equal amount in phase 2, as the figure shows, the amount of drop in both critic's value and actor's preference is equal. While this drop is sufficient for the critic's value, $V(s, a)$, to converge to r_{sh}, it is not enough to make the preference, $P(a \mid s)$, negative. Because the associated reward to action b is 0, its value and preference always remain 0. Hence, in phase 2, while the value of action a falls below the value of action b, the preference toward action a is still above the preference of action b. When the critic's value converges to r_{sh}, the prediction error by performing a converges to 0, and so no change in the value and preference associated with a occurs. The parameters of the simulation are detailed in the Supplementary Table 1, available online at http://www.mitpressjournals.org/doi/suppl/10.1162/NECO_a_00009.

value learning and preference adaptation are done using the same error signal, they remain coupled during learning. However, the preference for an appetitive stimulus grows more than its value. Hence, the punishment cannot cancel out drug-seeking behavior even when an acute punishment is substituted for the drug. It is also notable that the discrepancy between the value and the preference increases progressively until the prediction error becomes 0, and so the punishment that was previously able to cancel out pressing the level, is no longer able to stop drug taking after more experience with the drug.

These explanations are demonstrated in the following sections through simulating the model.

3.2 Compulsive Drug Seeking and Taking. There are some animal models for assessing compulsive behavior (Deroche-Gamonet et al., 2004; Vanderschuren & Everitt, 2004). We examine the behavior of the model in a procedure analogous to one of them carried out by Deroche-Gamonet and colleagues, which we refer to as Deroche-Gamonet's task. Their experiment has two stages. In the first stage, they used a self-administration procedure, a common method in animal models of addiction. Rats learned to freely obtain the drug by nose poking into a hole. In the second stage, the infusion of drug was associated with a shock punishment. The shock was signaled by a light stimulus that turned on with nose poking and off after shock delivery.

To simulate Deroche-Gamonet's task, we take into account an environment with two cues: a hole (that when available, the animal can obtain reward by nose-poking) and a light stimulus. The simulation consists of two steps. In the first step, the hole is available, but the light is not presented. The model can choose between nose poking and no action. Only by choosing nose poking does the animal obtain the drug. In the second step, the hole is available, and the light is presented. In this step, choosing nose poking results in the delivery of a shock punishment. The magnitude of the shock punishment is larger than the magnitude of the drug.

Results derived from simulating the model capture the essence of the behavioral observations mentioned above. Since the development of addiction depends on both the duration of drug exposure and individual vulnerability, we have decoupled these two influencing factors and examined them in two separate simulations.

Figure 2a shows the effects of drug exposure on the development of compulsive behavior. In this simulation, κ_c is initialized with a low value in order to fulfill vulnerability prerequisite. As Figure 2a shows, shock can suppress drug acquisition only after limited use. Figure 2b shows the impacts of individual vulnerability on the development of addiction-like behavior. As discussed in the preceding sections, vulnerability in our model is defined by lower-than-normal levels of κ_c—$\kappa_c < 1$. Different initial values of κ_c are simulated in Deroche-Gamonet's task, where the time horizon of

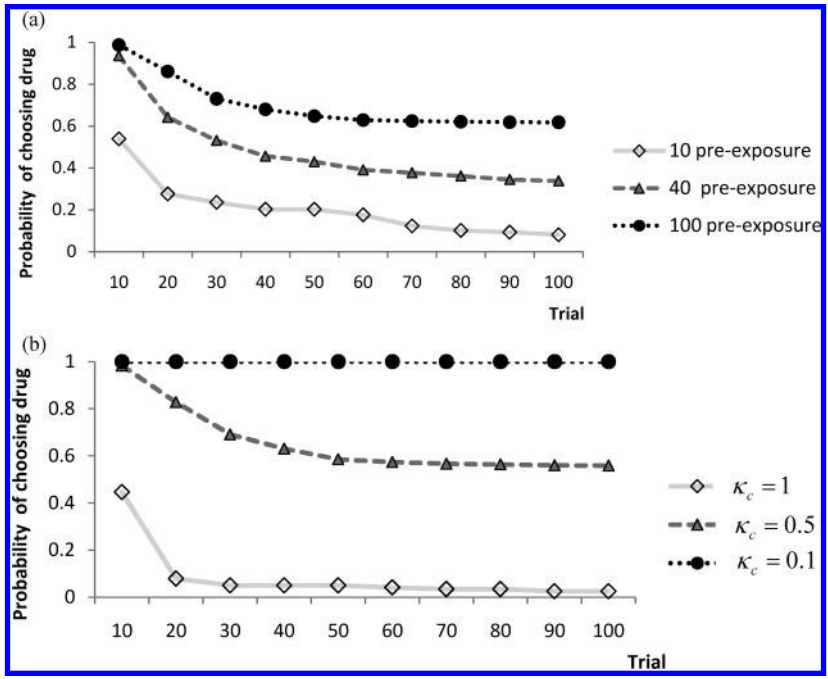

Figure 2: (a) The effect of duration of drug exposure on a vulnerable model with $\kappa_c = 0.5$ in Deroche-Gamonet's task. The performance of the model after 10, 40, and 100 trials of exposure to the drug in the face of shock punishment is depicted. Only after limited use does shock suppress drug acquisition. (b) The effect of individual vulnerability in Deroche-Gamonet's task. Three models with different κ_c self-administer drug equally (100 trials). The performance of these models in the face of shock punishment is shown. The development of compulsive behavior is inversely related to κ_c. That is, models with lower levels of receptors are more vulnerable to development of compulsive behavior. The parameters are detailed in Supplementary Table 2, available online at http://www.mitpressjournals.org/doi/suppl/10.1162/NECO_a_00009.

experiencing the drug in the training phase remains fixed. After the training phase, the behavior of the model for each initial value of κ_c is evaluated. Figure 2b shows that propensity to nose poking is inversely correlated with the magnitude of κ_c. Only for the models with $\kappa_c < 1$ does the preference for drug use outweigh its value. The lower the value of κ_c, the more the preference for the drug will be exaggerated, and hence, it is expected that low values of κ_c predict drug-seeking behavior, as shown in the figure.

3.3 Involvement of the Dorsal Striatum in Developing Addiction. A wealth of evidence supports the hypothesis that the involvement of the

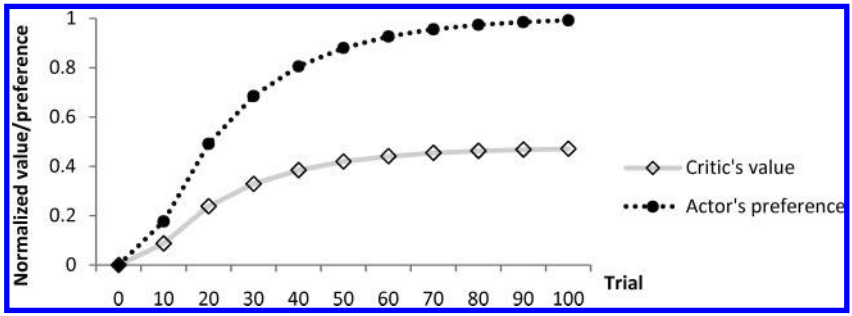

Figure 3: The value and preference associated with drug-seeking action in the first stage of Deroche-Gamonet's task for a vulnerable model with $\kappa_c = 0.5$. The model self-administers a drug for 100 trials. The actor's preference identifies propensity to drug use. The normalized value in the critic in the face of normalized preference (habit) in the actor with drug self-administration is depicted. The more experience with the drug, the more consolidated habit is in the actor. Due to the increasing discrepancy between the critic's value and the actor's preference, after long-term drug use, the critic is no longer able to direct the actor toward an optimal decision. The involvement of the actor in addiction is progressively augmented. The parameters are those used in the previous simulation and are detailed in Supplementary Table 2, available online at http://www.mitpressjournals.org/doi/suppl/10.1162/NECO_a_00009.

dorsal striatum is progressively augmented by the development of addiction. Since in our model the actor plays the role of the dorsal striatum, we evaluate the actor's preferences when the model self-administers the drug.

Assume a simple environment where the model takes drugs by pressing a lever. κ_c is initialized with a relatively low value in order to fulfill the vulnerability prerequisite. We assess the values and preferences for drug taking during the time of drug use. As Figure 3 shows, preference toward taking drugs is abnormally consolidated after chronic consumption of drugs by vulnerable individuals. Indeed, a punishment that was previously able to suppress drug seeking can no longer cancel out the drug-seeking habit. Hence, the involvement of the actor that models the dorsal striatal function is progressively augmented. In fact, through the development of addiction, the discrepancy between the value and the preference progressively increases, and so, by giving a punishment, the prediction error calculated by the critic is no longer able to direct the actor toward the optimal decision. This is because the maladaptive habit is consolidated abnormally in the actor. This behavior of the model is consistent with evidence showing that the dorsal striatum plays a critical role in habitual and compulsive drug seeking (Everitt & Robbins, 2005), and studies have shown that the metabolic activity of dorsal domains of striatum is progressively augmented (Porrino et al., 2004).

Two points are notable. First, although the actor's habitual response is abnormally consolidated toward taking the drug, it is the critic's deficit that is the source of this abnormal behavior. At the neural level, the model predicts that although the dorsal striatum plays the prominent role in maladaptive response, it is the NAc deficit that consolidates this maladaptive behavior. Second, only chronic, not acute, exposure to drugs increases the actor's preference. This is because the effects of the drug on κ_c appear only after chronic drug consumption.

3.4 Food Addiction and Pathological Gambling. Some studies have demonstrated that there is a common mechanism at least partly underlying obesity and addiction (Volkow et al., 2008; Volkow & Wise, 2005) and also pathological gambling and addiction (Potenza, 2008). For example, the role of the dopaminergic reward circuit in these diseases is prominent, and all of them occur in individuals with similar vulnerabilities in dopamine receptors (Steeves et al., 2009; Volkow et al., 2008). Regarding that all behaviors of the model depend on only high dopamine release in the NAc of a vulnerable individual (i.e., an individual with low availability of dopamine receptors in the NAc), the model can explain addiction-like behavior for foods or gambling, both of which, naturally produce dopamine in the striatum. Indeed, the euphoria produced by sex, food, or gambling makes them highly rewarding, and they produce a relatively high dopamine signal that can diminish the availability of NAc dopamine receptors, κ_c, in the vulnerable individual. Hence, although these rewards have no pharmacological effect, they can lead to an imbalance between learning the appetitive and aversive consequences of stimuli and lead vulnerable individuals to compulsive behavior.

The model thus predicts that experiencing addiction to any of these reinforcers will make the individual more vulnerable to addiction to the others. This is due to the diminishing effect of the first addiction-induced reinforcer on κ_c.

3.5 Parkinsonism and Its Relation to Addiction. Parkinson's disease (PD) is associated with a deficit in dorsal striatal dopamine and its dopamine afferents projecting from SNc. Although PD is known as a movement disorder, patients have some deficits in cognitive functions, especially those related to learning from appetitive and aversive outcomes. Importantly, PD patients display enhanced harm-avoidance behavior, that is, enhanced ability to avoid negative a outcome (Frank et al., 2004). One medication for these patients is D2 agonist, which improves the ability of patients to learn from positive outcomes, although it surprisingly exacerbates the performance in some tasks that require learning from negative outcome (Frank et al., 2004). Another puzzling observation is that while nonmedicated PD patients show fewer tendencies for smoking cigarettes or other forms of addiction such as pathological gambling, after medication with D2 agonist,

a subpopulation of them had an incidence of pathological gambling as high as 8%. This is surprising because evidence shows less than 1% of pathological gambling in general population (Dagher & Robbins, 2009). In this section, generalizing our approach in modeling NAc dopamine receptors to dorsal striatal receptors, we seek to explain the puzzling relationship between PD and addictive behaviors.

Frank and colleagues have hypothesized that D1 and D2 receptors play distinctive roles in learning from positive and negative outcomes, especially in more dorsal regions (Frank et al., 2004; Hazy et al., 2007). According to this hypothesis, D1 and D2 receptors are involved in learning from positive and negative prediction error, respectively. We symbolize the effect of D1 by κ_a^+, and D2 by κ_a^-. Similar to our approach in modeling the effects of receptors in the NAc (critic), we model receptors in the dorsal striatum (actor) as modulators of learning rate. We assume that the learning rate in the actor is $\alpha\kappa_a^+$ and $\alpha\kappa_a^-$ for positive and negative prediction error, respectively. In other words, taking into account the effects of dorsal receptors, the actor's preferences are updated by $\alpha\kappa_a^+\delta$ and $\alpha\kappa_a^-\delta$ for positive and negative feedback, respectively. Since PD is associated with a deficit in dorsal striatal dopamine system, we model PD by deficits in the actor. Here, in line with Frank and colleagues (Frank et al., 2007), we assume that off-medication PD patients have higher learning rates from negative outcomes compared to positive outcomes, that is, $\kappa_a^- > 1$. Due to $\kappa_a^- > 1$, the avoidance from negative outcomes is exaggerated in the actor during leaning.

This approach in modeling can explain why a Parkinsonian personality contrasts with an addiction personality in terms of balancing appetitive and aversive aspects of reinforcers. Assume a PD patient model (in which $\kappa_a^- > 1$) performs the task mentioned in Figure 1a. In the first phase, the value of pressing lever changes from 0 to the value of drug reward, r_d, and so the prediction error is always positive. In this phase, the critic's value for drug seeking is updated by $\kappa_c\delta$, and the actor's preference is updated by $\kappa_a^+\delta$. Since $\kappa_c = \kappa_a^+ = 1$, both the critic's value and the actor's preference for lever pressing converge to r_d. In the second phase, when a punishment r_{sh}, is substituted for the reward, r_{sh}, the value of drug seeking changes from r_d to r_{sh}, and so the prediction error is negative during this phase of learning. Hence, while the value of lever pressing is updated by $\alpha\delta$, the actor's preference for this action is updated by $\alpha\kappa_a^-\delta$. Since $\kappa_a^- > 1$, the avoidance exaggerates in the actor compared to the critic.

Consequently, we expect that PD patients, do not tend to different forms of addiction, because in these patients a negative outcome in the actor is exaggerated. This tendency broadly contrasts with the characteristic of individuals with a tendency to addiction: whereas Parkinsonism exaggerates negative outcomes in the actor, addiction vulnerability exaggerates appetitive outcomes. This explains why PD patients have a lower tendency to addiction-like behaviors (Dagher & Robbins, 2009).

Now we model the effect of D2 agonist, which is a medication for PD. In line with Frank et al. (2004, 2007), we assume that D2 agonist medication decreases the learning rate from a negative outcome in the dorsal striatum, κ_a^-. On the other hand, by stimulation of D2 receptors in the NAc, the subsequent dopamine release in this region increases. Hence, for PD patients with a comorbidity of addiction ($\kappa_c < 1$), the increase in dopamine release results in a reduction of dopamine receptors according to equations 2.8 and 2.9. Therefore, the medication not only reduces the avoidance form of negative outcome in the actor but also enhances a preference toward appetitive stimuli in PD patients with a comorbidity for addiction. This may explain why pathological gambling in PD patients medicated with a D2 agonist is reported to be higher than the general population. Indeed, this observation can be related to a reduction of κ_c in PD patients even with a limited vulnerability for addiction, which have κ_c slightly less than 1. According to equation 2.8, for such a vulnerable individual, a very high dopamine release is needed to reduce κ_c that might not be produced by gambling. Nonetheless, D2 agonist medication can reduce κ_c, and so a previously less vulnerable individual becomes highly vulnerable to the development of addiction.

After stopping treatment with a D2 agonist, the availability of D2 receptors in the NAc and the dorsal striatum returns to its initial levels, and thus the symptoms of addiction disappear. However, because levodopa is a drug that increases dopamine in striatum, equations 2.8 and 2.9 do not allow the dopamine receptors to recover, and so, by reducing D2 agonist dose contaminant with an increase in levodopa dose, the addictive symptoms persist.

It worth mentioning that although a negative outcome ($\delta < 0$) and aversive stimuli are different ($r < 0$) in nature, most tasks in the literature on PD that investigates deficit in learning are not able to discriminate these two different conditions. Further studies are needed to determine that a distinction reported in learning from positive versus negative prediction error is actually related to the prediction error itself or the valence of the stimuli (but see Frank et al., 2009; Moustafa, Cohen, Sherman, & Frank, 2008; Peterson et al., 2009). Importantly, Matsumoto and Hikosaka in a nonhuman primate study have reported that some SNc dopamine neurons excite by punishment-predicting stimuli (Matsumoto & Hikosaka, 2009; but see also Frank & Surmeier, 2009).

In the light of our model, a higher learning rate for appetitive stimuli in the critic in the face of other normal learning rates leads to behaviors similar to PD patients' behaviors. It is because the actual value of an appetitive stimulus will be attenuated in the actor. Hence, the model predicts a lower tendency to addiction-like behavior for any pathology that includes a higher level of both D1 and D2 dopamine receptors in the NAc but a normal level of dopamine receptors in the dorsal striatum.

4 Discussion

We proposed a computational framework for the hypothesis that individual differences in the NAc dopamine receptors predispose addiction-like behavior. The proposed model explains and simulates the compulsive behavior as a function of two factors: individual vulnerability and duration of drug exposure. We computationally elucidated how possibly the involvement of the dorsal striatum is progressively augmented and why abnormal habitual drug seeking and drug taking mainly depend on the dorsal striatum. The model proposes a common framework for describing addiction to natural rewards and addictive drugs. Finally, the proposed model explains how PD medication with D2 agonists might trigger pathological gambling, as well as other forms of addiction.

4.1 Individual Differences and Computational Models of Addiction. Previously, individual susceptibility to drug addiction was not addressed explicitly by abstract computational (Dayan, 2009; Dezfouli et al., 2009; Redish, 2004; Redish et al., 2007; Redish & Johnson, 2007; Zhang et al., 2009) or multilevel circuit models of addiction (Gutkin et al., 2006). Although the development of addiction in the previous models did not depend on individual differences systematically, we and others assumed that the free parameters of models can be different across individuals and this can lead to different patterns for the development of addiction. Our approach here is consistent with those studies; the parameter here is the learning rate. However, we have shown how the development of addiction can depend on this parameter. Moreover, we made a clear link between neural evidence and the parameter-mediated addiction in the proposed model.

Among the previous models, our approach has some similarities to that proposed by Gutkin and colleagues (2006), which is a neuronal network dynamical model for nicotine addiction. The important similarity is that both models explain addiction by interpreting the effect of drug on learning rates. In their model, a slow opponent process plays a critical role in drug addiction. It is assumed that a dopamine signal governs the gating of memory. On the other hand, long-term exposure to nicotine causes the dopamine signal to fall below a certain threshold needed for efficient learning. The model explains decreased harm avoidance based on an impaired learning mechanism. After long-term drug abuse, the model is unable to learn that drug seeking and taking is followed by harmful consequences. Considering the behavior of the model, it learns a nondrug reinforcer more slowly after long-term drug consumption, but after being learned, the behavior of the model does not differ from what it was before the chronic drug abuse. Compared to our model, it is more concrete and explains the process at a neuronal level. Nevertheless, although we model the effects of drugs on the critic's learning rate for appetitive stimuli, the results do not

depend on the slow learning speed; they depend instead on the imbalance between learning rates for appetitive and aversive stimuli in the critic.

It is also important to note that the lower availability of NAc dopamine receptors seems not to be the sole reason that individuals become susceptible to addiction; our proposed framework—the imbalance between appetitive and aversive learning in the critic and the actor—is not the only way to addiction. As explained in section 1, the interaction of drug and dopaminergic circuitry is multidimensional. The effects of drugs on other components of dopaminergic circuitry, such as phasic dopamine, can also lead to addiction (Dezfouli et al., 2009; Redish, 2004). Notably, the effects of drugs are not only restricted to the dopamine circuit (Ahmed, 2004); addiction can be the result of vulnerabilities in different subsystems involved in the decision process (Redish et al., 2008).

4.2 The Model Predicts Accelerated Habit Formation After Long-Term Drug Use. A clear prediction of the model is that drug use accelerates habit formation in instrumental tasks. In other words, by long-term drug abuse, due to the diminishing trend of κ_c, the discrepancy between the value of and the preference for obtaining rewards increases, even for stimuli that are not able to decrease dopamine receptors. Therefore, by substituting the reward with punishment (devaluation), due to the abnormal consolidated habit-like responses in the actor, the model continues to seek the natural reward for a longer-than-expected period. This prediction is in harmony with studies showing that animals exposed to cocaine exhibit more rigid and inflexible behavior than control animals when the predicted food is devalued (Nelson & Killcross, 2006; Schoenbaum & Setlow, 2005) or after reversal (Jentsch, Olausson, De La Garza, & Taylor, 2002; Schoenbaum, Saddoris, Ramus, Shaham, & Setlow, 2004; Takahashi, Roesch, Stalnaker, & Schoenbaum, 2007; Takahashi et al., 2008).

4.3 The First Leg of the Spiral Plays the Most Important Role in Our Model. Recent studies have revealed the importance of striatal-midbrain spiraling network in connecting ventral regions of the striatum to the dorsal regions. This network consists of a cascading serial connectivity that links the NAc shell to the more posterior VTA, which then projects to the NAc core and also to more dorsal regions of striatal organization, which in turn projects to SNc, and so on (Haber, Fudge, & McFarland, 2000; Ikemoto, 2007). Using an intrastriatal disconnection procedure, it has been shown that the striato-midbrain-striatal serial dopaminergic connectivity is essential for enhancing drug-seeking habits (Belin & Everitt, 2008). Our model reveals the importance of the two legs of this spiral computationally: posteromedial VTA to NAc shell and SNc to dorsal striatal projections.

Studies suggest a pivotal role for synaptic plasticity in the VTA that appears just 4 hours after a single exposure to cocaine and initiates plasticity in the NAc (Kauer & Malenka, 2007). Anatomically, this initiation

can be attributed to the first leg of the spiral. Our model reveals that the involvement of the first leg of the spiral in response to rewarding stimuli is important to initiate plasticity in the NAc. Hence, we suggest that early adaptations in the VTA might include increased sensitivity of its dopamine neurons to rewarding stimuli, and this hypersensitivity initiates adaptations in the NAc. In other words, due to the earlier actions of the drug on the VTA, more neurons in the VTA respond to appetitive but not aversive stimuli. This suggestion reveals the importance of studies that measure the responsiveness of the VTA dopamine neurons to rewarding stimuli in pre- and post-drug-taking phases. The first target for such studies can be the dopamine neurons that are excited by aversive stimuli and neurons that do not excite or inhibit by aversive stimuli (unresponsive neurons) (Brischoux, Chakraborty, Brierley, & Ungless, 2009).

4.4 Modeling Other Limbic Structures Is the Next Step to Complete the Proposed Model. The adaptations in the VTA can be mediated by mGluR1 receptors on the dopamine neurons (Mameli et al., 2009). The glutamatergic plasticity in the dopamine system is of great importance according to a wealth of evidence demonstrating that glutamatergic progression plays a critical role in reducing the ability of the brain's limbic structures to control behavior. This loss of control comes with overcoming the motor structures, such as the dorsal striatum and SNc, on the limbic structures, such as the PFC, the NAc, and the VTA (Kalivas, 2009). Nevertheless, modeling the neuroadaptations in the connections between PFC and NAc requires a computational key that normally mediates behavioral shift, which is important in particular for modeling relapse (Kalivas & O'Brien, 2008; Redish et al., 2007) and craving (Conrad et al., 2008; Redish & Johnson, 2007). Bayesian uncertainty of values estimated by these systems has been proposed as the computational key to switch behavior from goal-directed action to habitual response (Daw, Niv, & Dayan, 2005). At the neural level, the dopamine system and its interaction with the glutamate system might play an important role in behavioral shifting. Consistently, it has been shown that a gene in the PFC controlling prefrontal dopamine function is associated with uncertainty-based exploration (Frank et al., 2009). Also, D1 and D2 receptors in the NAc affect different facets of goal-directed behavior by modulating selectively synaptic input from hippocampus and medial PFC (Goto & Grace, 2005).

On the other hand, it seems that OFC, as a part of PFC, plays an important role in different phases of addiction. Since it is hypothesized that OFC encodes the incentive value of different choices (Schoenbaum, Roesch, Stalnaker, & Takahashi, 2009), modeling the role of this region can be the next step in advancing computational models of addiction. Specifically, it seems that OFC and its connection with NAc play an important role in impulsive behavior, which has a direct correlation with the tendency to use drugs (Roesch, Takahashi, Gugsa, Bissonette, & Schoenbaum, 2007). For example, a low density of D2 receptors in the NAc correlates with

impulsive behavior and impulsivity in turn predates compulsivity in rats (Belin et al., 2008; Dalley et al., 2007). Therefore, explanation of reported pre- and postdrug impulsivity in vulnerable individuals is an important feature of addiction that future models can focus on.

Acknowledgments

We are very grateful to Serge H. Ahmed for helpful discussions and also Samira Zaliasl and Fatemeh Fakhraie for comments on an early version of the manuscript. We are also very grateful to Hamed Ekhtiari who helped us a great deal during this research. Also, we acknowledge useful comments from the anonymous reviewers who helped to greatly improve the letter.

References

Ahmed, S. H. (2004). Addiction as compulsive reward prediction. *Science*, *306*(5703), 1901–1902.

Ahmed, S. H. (2008). The origin of addictions by means of unnatural decision. *Behavioral and Brain Sciences*, *31*(4), 437–438.

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. (2005). Transition to drug addiction: A negative reinforcement model based on an allostatic decrease in reward function. *Psychopharmacology*, *180*(3), 473–490.

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

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), 244–268.

Atallah, H. E., Lopez-Paniagua, D., Rudy, J. W., & O'Reilly, R. C. (2007). Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nature Neuroscience*, *10*(1), 126–131.

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

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*, *27*(31), 8161–8165.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141.

Belin, D., & Everitt, B. J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron*, *57*(3), 432–441.

Belin, D., Mar, A. C., Dalley, J. W., Robbins, T. W., & Everitt, B. J. (2008). High impulsivity predicts the switch to compulsive cocaine-taking. *Science*, *320*(5881), 1352–1355.

Berke, J. D., & Hyman, S. E. (2000). Addiction, dopamine, and the molecular mechanisms of memory. *Neuron*, *25*(3), 515–532.

Beveridge, T. J. R., Smith, H. R., Nader, M. A., & Porrino, L. J. (2009). Abstinence from chronic cocaine self-administration alters striatal dopamine systems in rhesus monkeys. *Neuropsychopharmacology*, *34*(5), 1162–1171.

Bolanos, C. A., Perrotti, L. I., Edwards, S., Eisch, A. J., Barrot, M., Olson, V. G., et al. (2003). Phospholipase C γ in distinct regions of the ventral tegmental area differentially modulates mood-related behaviors. *J. Neurosci.*, *23*(20), 7569–7576.

Brischoux, F., Chakraborty, S., Brierley, D. I., & Ungless, M. A. (2009). Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(12), 4894–4899.

Caine, S. B., Negus, S. S., Mello, N. K., & Bergman, J. (1999). Effects of dopamine D(1-like) and D(2-like) agonists in rats that self-administer cocaine. *Journal of Pharmacology and Experimental Therapeutics*, *291*(1), 353–360.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, *26*(3), 321–352.

Carelli, R. M. (2002). Nucleus accumbens cell firing during goal-directed behaviors for cocaine vs. "natural" reinforcement. *Physiology and Behavior*, *76*(3), 379–387.

Carlezon, W. A., Boundy, V. A., Haile, C. N., Lane, S. B., Kalb, R. G., Neve, R. L., et al. (1997). Sensitization to morphine induced by viral-mediated gene transfer. *Science*, *277*(5327), 812–815.

Carlezon, W. A., Devine, D. P., & Wise, R. A. (1995). Habit-forming actions of nomifensine in nucleus accumbens. *Psychopharmacology*, *122*(2), 194–197.

Carlezon, W. A., Haile, C. N., Coppersmith, R., Hayashi, Y., Malinow, R., Neve, R. L., et al. (2000). Distinct sites of opiate reward and aversion within the midbrain identified using a herpes simplex virus vector expressing GluR1. *J. Neurosci.*, *20*(5), 62RC.

Conrad, K. L., Tseng, K. Y., Uejima, J. L., Reimers, J. M., Heng, L., Shaham, Y., et al. (2008). Formation of accumbens GluR2-lacking AMPA receptors mediates incubation of cocaine craving. *Nature*, *454*(7200), 118–121.

Corrigall, W. A., & Coen, K. M. (1991). Cocaine self-administration is increased by both D1 and D2 dopamine antagonists. *Pharmacology, Biochemistry, and Behavior*, *39*(3), 799–802.

Dagher, A., & Robbins, T. W. (2009). Personality, addiction, dopamine: Insights from Parkinson's disease. *Neuron*, *61*(4), 502–510.

Dalley, J. W., Fryer, T. D., Brichard, L., Robinson, E. S. J., Theobald, D. E. H., Lääne, et al. (2007). Nucleus accumbens D2/3 receptors predict trait impulsivity and cocaine reinforcement. *Science*, *315*(5816), 1267–1270.

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*(4–6), 603–616.

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.

Dayan, P. (2009). Dopamine, reinforcement learning, and addiction. *Pharmacopsychiatry*, *42* (Suppl. 1), S56–65.

Dayan, P., & Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. *PLoS Computational Biology*, *4*(2), e4.

Deroche-Gamonet, V., Belin, D., & Piazza, P. V. (2004). Evidence for addiction-like behavior in the rat. *Science*, *305*(5686), 1014–1017.

de Wit, H., Uhlenhuth, E. H., & Johanson, C. E. (1986). Individual differences in the reinforcing and subjective effects of amphetamine and diazepam. *Drug and Alcohol Dependence*, *16*(4), 341–360.

Dezfouli, A., Piray, P., Keramati, M. M., Ekhtiari, H., Lucas, C., & Mokri, A. (2009). A Neurocomputational Model for Cocaine Addiction. *Neural Computation*, *21*, 2869–2893.

Di Chiara, G., Bassareo, V., Fenu, S., De Luca, M. A., Spina, L., Cadoni, C., et al. (2004). Dopamine and drug addiction: The nucleus accumbens shell connection. *Neuropharmacology*, *47* (Suppl. 1), 227–241.

Edwards, S., Whisler, K. N., Fuller, D. C., Orsulak, P. J., & Self, D. W. (2007). Addiction-related alterations in D1 and D2 dopamine receptor behavioral responses following chronic cocaine self-administration. *Neuropsychopharmacology*, *32*(2), 354–366.

Everitt, B. J., Belin, D., Economidou, D., Pelloux, Y., Dalley, J. W., & Robbins, T. W. (2008). Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1507), 3125–3135.

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.

Fenu, S., Spina, L., Rivas, E., Longoni, R., & Di Chiara, G. (2006). Morphine-conditioned single-trial place preference: Role of nucleus accumbens shell dopamine receptors in acquisition, but not expression. *Psychopharmacology*, *187*(2), 143–153.

Ford, C. P., Mark, G. P., & Williams, J. T. (2006). Properties and opioid inhibition of mesolimbic dopamine neurons vary according to target location. *J. Neurosci.*, *26*(10), 2788–2797.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*(8), 1062–1068.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(41), 16311–16316.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*, *306*(5703), 1940–1943.

Frank, M. J., & Surmeier, D. J. (2009). Do substantia nigra dopaminergic neurons differentiate between reward and punishment? *Journal of Molecular Cell Biology*, *1*(1), 15–16.

Garavan, H., Pankiewicz, J., Bloom, A., Cho, J. K., Sperry, L., Ross, T. J., et al. (2000). Cue-induced cocaine craving: Neuroanatomical specificity for drug users and drug stimuli. *American Journal of Psychiatry*, *157*(11), 1789–1798.

Goldstein, R. Z., Alia-Klein, N., Tomasi, D., Zhang, L., Cottone, L. A., Maloney, T., et al. (2007). Is decreased prefrontal cortical sensitivity to monetary reward associated with impaired motivation and self-control in cocaine addiction? *American Journal of Psychiatry*, *164*(1), 43–51.

Goto, Y., & Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature Neuroscience*, *8*(6), 805–812.

Gutkin, B. S., Dehaene, S., & Changeux, J. (2006). A neurocomputational hypothesis for nicotine addiction. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 1106–1111.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J. Neurosci.*, *20*(6), 2369–2382.

Haber, S. N., & Knutson, B. (2009). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*, 4–26.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1485), 1601–1613.

Hopf, F. W., Cascini, M. G., Gordon, A. S., Diamond, I., & Bonci, A. (2003). Cooperative activation of dopamine D1 and D2 receptors increases spike firing of nucleus accumbens neurons via G-protein betagamma subunits. *Journal of Neuroscience*, *23*(12), 5079–5087.

Hyman, S. E. (2005). Addiction: A disease of learning and memory. *American Journal of Psychiatry*, *162*(8), 1414–1422.

Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). Neural mechanisms of addiction: The role of reward-related learning and memory. *Annual Review of Neuroscience*, *29*, 565–598.

Ikemoto, S. (2003). Involvement of the olfactory tubercle in cocaine reward: Intracranial self-administration studies. *J. Neurosci.*, *23*(28), 9305–9311.

Ikemoto, S. (2007). Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Research Reviews*, *56*(1), 27–78.

Ikemoto, S., Glazier, B. S., Murphy, J. M., & McBride, W. J. (1997). Role of dopamine D1 and D2 receptors in the nucleus accumbens in mediating reward. *J. Neurosci.*, *17*(21), 8580–8587.

Ikemoto, S., Murphy, J. M., & McBride, W. J. (1998). Regional differences within the rat ventral tegmental area for muscimol self-infusions. *Pharmacology, Biochemistry, and Behavior*, *61*(1), 87–92.

Ikemoto, S., Qin, M., & Liu, Z. (2005). The functional divide for primary reinforcement of D-amphetamine lies between the medial and lateral ventral striatum: Is the division of the accumbens core, shell, and olfactory tubercle valid? *Journal of Neuroscience*, *25*(20), 5061–5065.

Ikemoto, S., & Wise, R. A. (2002). Rewarding effects of the cholinergic Agents carbachol and neostigmine in the posterior ventral tegmental area. *J. Neurosci.*, *22*(22), 9895–9904.

Inoue, Y., Yao, L., Hopf, F. W., Fan, P., Jiang, Z., Bonci, A., et al. (2007). Nicotine and ethanol activate protein kinase a synergistically via Gi beta γ subunits in nucleus accumbens/ventral tegmental cocultures: The role of dopamine D1/D2 and adenosine A2A receptors. *J. Pharmacol. Exp. Ther.*, *322*(1), 23–29.

Ito, M., & Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *Journal of Neuroscience*, *29*(31), 9861–9874.

Ito, R., Dalley, J. W., Robbins, T. W., & Everitt, B. J. (2002). Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. *Journal of Neuroscience*, *22*(14), 6247–6253.

Janak, P. H., Chen, M., & Caulder, T. (2004). Dynamics of neural coding in the accumbens during extinction and reinstatement of rewarded behavior. *Behavioural Brain Research*, *154*(1), 125–135.

Jentsch, J. D., Olausson, P., De La Garza, R., & Taylor, J. R. (2002). Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. *Neuropsychopharmacology*, *26*(2), 183–190.

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

Kalivas, P. W., & O'Brien, C. P. (2008). Drug addiction as a pathology of staged neuroplasticity. *Neuropsychopharmacology*, *33*(1), 166–180.

Kauer, J. A., & Malenka, R. C. (2007). Synaptic plasticity and addiction. *Nature Reviews. Neuroscience*, *8*(11), 844–858.

Kim, H., Sul, J. H., Huh, N., Lee, D., & Jung, M. W. (2009). Role of striatum in updating values of chosen actions. *Journal of Neuroscience*, *29*(47), 14701–14712.

Kim, Y. B., Huh, N., Lee, H., Baeg, E. H., Lee, D., & Jung, M. W. (2007). Encoding of action history in the rat ventral striatum. *Journal of Neurophysiology*, *98*(6), 3548–3556.

Klein, T. A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D. Y., & Ullsperger, M. (2007). Genetically determined differences in learning from errors. *Science*, *318*(5856), 1642–1645.

Koob, G. F., & Le Moal, M. (2005). Plasticity of reward neurocircuitry and the "dark side" of drug addiction. *Nature Neuroscience*, *8*(11), 1442–1444.

Laruelle, M., D'Souza, C. D., Baldwin, R. M., Abi-Dargham, A., Kanes, S. J., Fingado, C. L., et al. (1997). Imaging D2 receptor occupancy by endogenous dopamine in humans. *Neuropsychopharmacology*, *17*(3), 162–174.

Lau, B., & Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. *Neuron*, *58*(3), 451–463.

Lauwereyns, J., Watanabe, K., Coe, B., & Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature*, *418*(6896), 413–417.

Mach, R. H., Nader, M. A., Ehrenkaufer, R. L., Line, S. W., Smith, C. R., Gage, H. D., et al. (1997). Use of positron emission tomography to study the dynamics of psychostimulant-induced dopamine release. *Pharmacology, Biochemistry, and Behavior*, *57*(3), 477–486.

Mameli, M., Halbout, B., Creton, C., Engblom, D., Parkitna, J. R., Spanagel, R., & Lüscher, C. (2009). Cocaine-evoked synaptic plasticity: Persistence in the VTA triggers adaptations in the NAc. *Nature Neuroscience*, *12*(8), 1036–1041.

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

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841.

McClure, S. M., Daw, N. D., & Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neurosciences*, *26*(8), 423–428.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*(7010), 760–767.

Moore, R. J., Vinsant, S. L., Nader, M. A., Porrino, L. J., & Friedman, D. P. (1998). Effect of cocaine self-administration on striatal dopamine D1 receptors in rhesus monkeys. *Synapse*, *28*(1), 1–9.

Morgan, D., Grant, K. A., Gage, H. D., Mach, R. H., Kaplan, J. R., Prioleau, O., et al. (2002). Social dominance in monkeys: Dopamine D2 receptors and cocaine self-administration. *Nature Neuroscience*, *5*(2), 169–174.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience*, *9*(8), 1057–1063.

Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in Parkinsonism. *Journal of Neuroscience*, *28*(47), 12294–12304.

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. Series B, Biological Sciences*, *363*(1507), 3223–3232.

Nader, M. A., Daunais, J. B., Moore, T., Nader, S. H., Moore, R. J., Smith, H. R., et al. (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., et al. (2006). PET imaging of dopamine D2 receptors during chronic cocaine self-administration in monkeys. *Nature Neuroscience*, *9*(8), 1050–1056.

Nelson, A., & Killcross, S. (2006). Amphetamine exposure enhances habit formation. *J. Neurosci.*, *26*(14), 3805–3812.

Nicola, S. M. (2007). The nucleus accumbens as part of a basal ganglia action selection circuit. *Psychopharmacology*, *191*(3), 521–550.

Noble, E. P. (2003). D2 dopamine receptor gene in psychiatric and neurologic disorders and its phenotypes. *American Journal of Medical Genetics, Part B, Neuropsychiatric Genetics*, *116B*(1), 103–125.

O'Brien, C. P., Ehrman, R., & Ternes, J. (1986). Classical conditioning in human opioid dependence. In S. R. Goldberg and I. P. Stolerman (Eds.), *Behavioral Analysis of Drug Dependence* (pp. 329–356). London: Academic Press.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–454.

Olson, V. G., Zabetian, C. P., Bolanos, C. A., Edwards, S., Barrot, M., Eisch, A. J., et al. (2005). Regulation of drug reward by cAMP response element-binding protein: Evidence for two functionally distinct subregions of the ventral tegmental area. *J. Neurosci.*, *25*(23), 5553–5562.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*, 563–593.

Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., et al. (2007). Shaping of motor responses by incentive values through the basal ganglia. *Journal of Neuroscience*, *27*(5), 1176–1183.

Peterson, D. A., Elliott, C., Song, D. D., Makeig, S., Sejnowski, T. J., & Poizner, H. (2009). Probabilistic reversal learning is impaired in Parkinson's disease. *Neuroscience*, *163*(4), 1092–1101.

Pezze, M., Dalley, J. W., & Robbins, T. W. (2007). Differential roles of dopamine D1 and D2 receptors in the nucleus accumbens in attentional performance on the five-choice serial reaction time task. *Neuropsychopharmacology*, *32*(2), 273–283.

Piazza, P. V., Deminière, J., Le Moal, M., & Simon, H. (1989). Factors that predict individual vulnerability to amphetamine self-administration. *Science*, *245*(4925), 1511–1513.

Porrino, L. J., Daunais, J. B., Smith, H. R., & Nader, M. A. (2004). The expanding effects of cocaine: Studies in a nonhuman primate model of cocaine self-administration. *Neuroscience and Biobehavioral Reviews*, *27*(8), 813–820.

Porrino, L. J., Lyons, D., Smith, H. R., Daunais, J. B., & Nader, M. A. (2004). Cocaine self-administration produces a progressive involvement of limbic, association, and sensorimotor striatal domains. *J. Neurosci.*, *24*(14), 3554–3562.

Potenza, M. N. (2008). The neurobiology of pathological gambling and drug addiction: An overview and new findings. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *363*(1507), 3181–3189.

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

Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: Vulnerabilities in the decision process. *Behavioral and Brain Sciences*, *31*(4), 415–437.

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.

Redish, A. D., & Johnson, A. (2007). A computational model of craving and obsession. *Annals of the New York Academy of Sciences*, *1104*, 324–339.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research*, *18*(3), 247–291.

Rodd, Z. A., Bell, R. L., Zhang, Y., Murphy, J. M., Goldstein, A., Zaffaroni, A., et al. (2005). Regional heterogeneity for the intracranial self-administration of ethanol and acetaldehyde within the ventral tegmental area of alcohol-preferring (P)

rats: Involvement of dopamine and serotonin. *Neuropsychopharmacology*, *30*(2), 330–338.

Rodd, Z. A., Melendez, R. I., Bell, R. L., Kuc, K. A., Zhang, Y., Murphy, J. M., et al. (2004). Intracranial self-administration of ethanol within the ventral tegmental area of male wistar rats: Evidence for involvement of dopamine neurons. *J. Neurosci.*, *24*(5), 1050–1057.

Rodd-Henricks, Z. A., McKinzie, D. L., Li, T., Murphy, J. M., & McBride, W. J. (2002). Cocaine is self-administered into the shell but not the core of the nucleus accumbens of Wistar rats. *Journal of Pharmacology and Experimental Therapeutics*, *303*(3), 1216–1226.

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, *10*(12), 1615–1624.

Roesch, M. R., Singh, T., Brown, P. L., Mullins, S. E., & Schoenbaum, G. (2009). Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *Journal of Neuroscience*, *29*(42), 13365–13376.

Roesch, M. R., Takahashi, Y., Gugsa, N., Bissonette, G. B., & Schoenbaum, G. (2007). Previous cocaine exposure makes rats hypersensitive to both delay and reward magnitude. *Journal of Neuroscience*, *27*(1), 245–250.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, *310*(5752), 1337–1340.

Schoenbaum, G., Roesch, M. R., Stalnaker, T. A., & Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews Neuroscience*, *10*(12), 885–892.

Schoenbaum, G., Saddoris, M. P., Ramus, S. J., Shaham, Y., & Setlow, B. (2004). Cocaine-experienced rats exhibit learning deficits in a task sensitive to orbitofrontal cortex lesions. *European Journal of Neuroscience*, *19*(7), 1997–2002.

Schoenbaum, G., & Setlow, B. (2005). Cocaine makes actions insensitive to outcomes but not extinction: Implications for altered orbitofrontal-amygdalar function. *Cerebral Cortex*, *15*(8), 1162–1169.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Self, D. W. (2004). Regulation of drug-taking and -seeking behaviors by neuroadaptations in the mesolimbic dopamine system. *Neuropharmacology*, *47* (Suppl. 1), 242–255.

Sellings, L. H. L., & Clarke, P. B. S. (2003). Segregation of amphetamine reward and locomotor stimulation between nucleus accumbens medial shell and core. *J. Neurosci.*, *23*(15), 6295–6303.

Sellings, L. H. L., McQuade, L. E., & Clarke, P. B. S. (2006a). Evidence for multiple sites within rat ventral striatum mediating cocaine-conditioned place preference and locomotor activation. *Journal of Pharmacology and Experimental Therapeutics*, *317*(3), 1178–1187.

Sellings, L. H. L., McQuade, L. E., & Clarke, P. B. S. (2006b). Characterization of dopamine-dependent rewarding and locomotor stimulant effects of intravenously-administered methylphenidate in rats. *Neuroscience*, *141*(3), 1457–1468.

Shin, R., Qin, M., Liu, Z., & Ikemoto, S. (2008). Intracranial self-administration of MDMA into the ventral striatum of the rat: Differential roles of the nucleus accumbens shell, core, and olfactory tubercle. *Psychopharmacology*, *198*(2), 261–270.

Spina, L., Fenu, S., Longoni, R., Rivas, E., & Di Chiara, G. (2006). Nicotine-conditioned single-trial place preference: Selective role of nucleus accumbens shell dopamine D1 receptors in acquisition. *Psychopharmacology*, *184*(3–4), 447–455.

Steeves, T. D. L., Miyasaki, J., Zurowski, M., Lang, A. E., Pellecchia, G., Van Eimeren, et al. (2009). Increased striatal dopamine release in Parkinsonian patients with pathological gambling: a [11C] raclopride PET study. *Brain*, *132*(Pt. 5), 1376–1385.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Taha, S. A., & Fields, H. L. (2005). Encoding of palatability and appetitive behaviors by distinct neuronal populations in the nucleus accumbens. *Journal of Neuroscience*, *25*(5), 1193–1202.

Takahashi, Y., Roesch, M. R., Stalnaker, T. A., & Schoenbaum, G. (2007). Cocaine exposure shifts the balance of associative encoding from ventral to dorsolateral striatum. *Frontiers in Integrative Neuroscience*, *1*, 11.

Takahashi, Y., Schoenbaum, G., & Niv, Y. (2008). Silencing the critics: Understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. *Frontiers in Neuroscience*, *2*(1), 86–99.

Thanos, P. K., Volkow, N. D., Freimuth, P., Umegaki, H., Ikari, H., Roth, G., et al. (2001). Overexpression of dopamine D2 receptors reduces alcohol self-administration. *Journal of Neurochemistry*, *78*(5), 1094–1103.

van der Meer, M. A. A., & Redish, A. D. (2009). Covert expectation-of-reward in rat ventral striatum at decision points. *Frontiers in Integrative Neuroscience*, *3*, 1.

Vanderschuren, L. J. M. J., Di Ciano, P., & Everitt, B. J. (2005). Involvement of the dorsal striatum in cue-controlled cocaine seeking. *Journal of Neuroscience*, *25*(38), 8665–8670.

Vanderschuren, L. J. M. J., & Everitt, B. J. (2004). Drug seeking becomes compulsive after prolonged cocaine self-administration. *Science*, *305*(5686), 1017–1019.

Volkow, N. D., Fowler, J. S., Wang, G., & Swanson, J. M. (2004). Dopamine in drug abuse and addiction: Results from imaging studies and treatment implications. *Molecular Psychiatry*, *9*(6), 557–569.

Volkow, N. D., Wang, G., Begleiter, H., Porjesz, B., Fowler, J. S., Telang, F., et al. (2006). High levels of dopamine D2 receptors in unaffected members of alcoholic families: Possible protective factors. *Archives of General Psychiatry*, *63*(9), 999–1008.

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

Volkow, N. D., Wang, G., Fowler, J. S., & Telang, F. (2008). Overlapping neuronal circuits in addiction and obesity: Evidence of systems pathology. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *363*(1507), 3191–3200.

Volkow, N. D., Wang, G., Telang, F., Fowler, J. S., Logan, J., Childress, A., et al. (2006). Cocaine cues and dopamine in dorsal striatum: Mechanism of craving in cocaine addiction. *Journal of Neuroscience*, *26*(24), 6583–6588.

Volkow, N. D., & Wise, R. A. (2005). How can drug addiction help us understand obesity? *Nature Neuroscience*, *8*(5), 555–560.

Wong, D. F., Kuwabara, H., Schretlen, D. J., Bonson, K. R., Zhou, Y., Nandi, A., et al. (2006). Increased occupancy of dopamine receptors in human striatum during cue-elicited cocaine craving. *Neuropsychopharmacology*, *31*(12), 2716–2727.

Zangen, A., Ikemoto, S., Zadina, J. E., & Wise, R. A. (2002). Rewarding and psychomotor stimulant effects of endomorphin-1: Anteroposterior differences within the ventral tegmental area and lack of effect in nucleus accumbens. *J. Neurosci.*, *22*(16), 7225–7233.

Zangen, A., Solinas, M., Ikemoto, S., Goldberg, S. R., & Wise, R. A. (2006). Two brain sites for cannabinoid reward. *Journal of Neuroscience*, *26*(18), 4901–4907.

Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S., & Aldridge, J. W. (2009). A neural computational model of incentive Salience. *PLoS Computational Biology*, *5*(7), e1000437.

Received August 20, 2009; accepted February 18, 2010.

This article has been cited by: