Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an Actor/Critic model

Yuji Takahashi[1], *Geoffrey Schoenbaum*[1] *and Yael Niv*[2,3,†]

[1] Department of Anatomy and Neurobiology, School of Medicine, University of Maryland, Baltimore, MD, USA.
[2] Psychology Department, Princeton University, Princeton, NJ, USA.
[3] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA.

Edited by:
Sidney A. Simon, Duke University, USA

Reviewed by:
Rui M. Costa, National Institutes of Health, USA
Bernard W. Balleine, University of California Los Angeles, USA

[†]***Correspondence:***

Yael Niv *will be Assistant Professor at the Princeton Neuroscience Institute and the Psychology Department at Princeton University beginning this fall. Currently a postdoctoral fellow at Princeton, she received her PhD from the Hebrew University in Jerusalem, after conducting her doctoral research there at the Interdisciplinary Center for Neural Computation and at the Gatsby Computational Neuroscience Unit at UCL, London. In her research, she strives to understand animal and human learning and decision-making on the computational, behavioral, and neural levels.*
yael@princeton.edu

A critical problem in daily decision making is how to choose actions now in order to bring about rewards later. Indeed, many of our actions have long-term consequences, and it is important to not be myopic in balancing the pros and cons of different options, but rather to take into account both immediate and delayed consequences of actions. Failures to do so may be manifest as persistent, maladaptive decision-making, one example of which is addiction where behavior seems to be driven by the immediate positive experiences with drugs, despite the delayed adverse consequences. A recent study by Takahashi et al. (2007) investigated the effects of cocaine sensitization on decision making in rats and showed that drug use resulted in altered representations in the ventral striatum and the dorsolateral striatum, areas that have been implicated in the neural instantiation of a computational solution to optimal long-term actions selection called the Actor/Critic framework. In this Focus article we discuss their results and offer a computational interpretation in terms of drug-induced impairments in the Critic. We first survey the different lines of evidence linking the subparts of the striatum to the Actor/Critic framework, and then suggest two possible scenarios of breakdown that are suggested by Takahashi et al.'s (2007) data. As both are compatible with the current data, we discuss their different predictions and how these could be empirically tested in order to further elucidate (and hopefully inch towards curing) the neural basis of drug addiction.

Keywords: Actor/Critic, cocaine, reinforcement learning, striatum

INTRODUCTION

A critical problem in animal and human decision making is how to choose behavior that will *in the long run* lead to reward. For instance, when playing a game of chess, early moves may be crucial to the long-term goal of winning: One set of moves may set the stage for a stunning victory; another result in a precipitous defeat. The difficulty in learning to play chess stems from the fact that the outcome, i.e., the win or loss at the end of the game, may be delayed with respect to such early actions. Moreover, once the game is over it is not clear, of the many actions taken throughout the game, which were the critical ones that should be learned for future games. Similar problems are encountered to a greater or lesser extent throughout our daily lives, in the many situations in which feedback for actions is not immediately available. Failure to solve this so-called "credit assignment" problem (Barto et al., 1983; Sutton and Barto, 1998) may be manifest as persistent, maladaptive decision-making: rather than being influenced by long-term goals, actions will be driven predominantly by proximal outcomes, thereby appearing

impulsive and "irrational", and succumbing easily to the temptations of short term pleasures. One possible example of this is addiction, where behavior seems to be driven by the early and immediate positive experiences with drugs, despite the often remote or delayed adverse consequences of continued drug use.

Although they do not play chess, rats encounter similar credit assignment problems in tasks that require multiple actions in order to obtain rewards or avoid punishments (e.g., when navigating a maze to a goal location or making a series of lever pressing and food-magazine approach actions in an operant chamber). Interestingly, exposure to addictive drugs has been shown to cause maladaptive decision-making in rats performing such tasks (Bechara and Damasio, 2002; Calu et al., 2007; Jentsch et al., 2002; Nelson and Killcross, 2006; Schoenbaum and Setlow, 2005; Schoenbaum et al., 2004; Simon et al., 2007; Vanderschuren and Everitt, 2004). In a study recently published in Frontiers in Integrative Neuroscience by Takahashi et al. (2007), these effects were linked to long-term changes in associative representations in two subdivisions of the striatum, the ventral striatum and the dorsolateral striatum. These areas have previously been implicated in the neural instantiation of a computational solution to the credit assignment problem, called the Actor/Critic learning and decision-making architecture.

In this Focus article, we discuss their results and offer a computational interpretation in terms of drug-induced modifications of the Actor/Critic model. We begin by describing the Actor/Critic framework, and presenting a short survey of the different lines of evidence linking it to the ventral and dorsolateral striatum and their dopaminergic afferents. We then explore the possible implications of the results reported by Takahashi et al. (2007) for such a model, by describing two possible scenarios in which the Actor must behave in the absence of a functional Critic. We conclude with directions of future research, which this study immediately suggests.

THE ACTOR/CRITIC MODEL: A PLAYER AND A COACH

One way to alleviate the credit assignment problem is by using as the advice of a coach: whether in sports or in chess, a coach who can give immediate feedback to actions even before the game has ended, can make all the difference in terms of learning. What the player really needs is someone that can tell her whether her current action is good or bad. Importantly, for this information to be useful, it should evaluate the action with respect to the long-term goal of winning: a coach that reinforces a locally good action of capturing a pawn, despite the fact that a more strategic action had to be forfeited for this capture to happen, is rather dispensable.

The Actor/Critic architecture (**Figure 1A**) does exactly this: the Actor chooses actions according to some policy of behavior, and the Critic offers immediate feedback that tells the Actor whether the action selected was good or bad form the point of view of obtaining rewards in the long run (Barto et al., 1983). The crux of this model is a simple learning rule based on a *Temporal Difference (TD) prediction error* (Barto et al., 1989; Sutton, 1988; Sutton and Barto, 1990):

$$\delta_t = r(S_t) + V(S_t) - V(S_{t-1}),$$

by which the Critic learns what feedback to give to each action, and the Actor learns an improved action selection policy. The components of this prediction error will be defined below.

In order to compute such a feedback prediction error, the Critic must first evaluate the present state of the world (S_t – the current situation, comprised of the available stimuli and context) in terms of the long-term expected sum of future rewards when commencing behavior from this state. That is, the Critic can be seen as having two roles[1]: one is to assign a value $V(S_t)$ to the current state, and the other is to combine state values with obtained rewards and compute the prediction error above. As we shall show below, these two roles are intimately intertwined.

In this reinforcement learning framework, state evaluations are essentially an estimation of the sum of future rewards from this point onward. As a result, they should obey a simple local consistency relation: the reward expected from this time point onward should be equal to the immediate reward at the next time point plus any rewards expected from that time point onward, that is $V(S_{t-1}) = r(S_t) + V(S_t)$ where $r(S_t)$ denotes the (possibly stochastic or 0) immediate reward at state S_t. This consistency forms the basis for the TD prediction error: δ_t above quantifies the inconsistency between successive values, i.e., the difference between reward predictions in consecutive states.

To learn correct values, all the Critic needs to do is to compute errors in its own value predictions

[1]Some texts include only the value learning component in the Critic, with the prediction error computation external to both Actor and Critic. We have chosen here to include the prediction error within the Critic for the sake of consistency of the terminology: as will become clear in the following text, it is the prediction error that criticizes the Actor, not the state values.

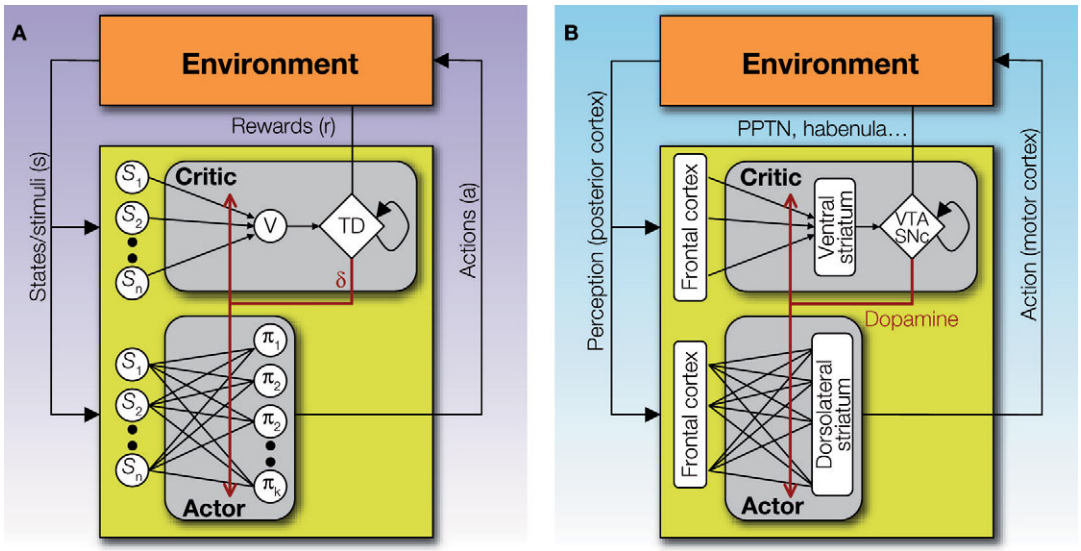

Figure 1 | **The basic Actor/Critic architecture and its suggested neural implementation.** ***(A)*** The (external or internal) environment provides two signals to the system: S, indicating the current state or stimuli, and r indicating the current reward. The Actor comprises of a mapping between states S and action propensities $\pi(a|S)$ (through modifiable weights or associative strengths). Its ultimate output is an action which then feeds back into the environment and serves to (possibly) earn rewards and change the state of the environment. The Critic comprises of a mapping between states S and values V (also through modifiable weights). The value of the current state provides input to a temporal difference (TD) module that integrates the value of the current state, the value of the previous state (indicated by the feedback arrow) and the current reward, to compute a prediction error signal $\delta_t = r(S_t) + V(S_t) - V(S_{t-1})$. This signal is used to modify the mappings in both the Actor and the Critic. ***(B)*** A suggested mapping of the Actor/Critic architecture onto neural substrates in the cortex and basal ganglia. The mapping between states and actions in the Actor is realized through plastic synapses between the cortex and the dorsolateral striatum. The mapping between states and their values is realized though similarly modifiable synaptic strengths in cortical projections to the ventral striatum. The prediction error is computed in the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) – the two midbrain dopaminergic nuclei – based on state values from ventral striatal afferents, and outcome information from sources such as the pedunculopontine nucleus (PPTN), the habenula etc. (Christoph et al., 1986; Ji and Shepard, 2007; Kobayashi and Okada, 2007; Matsumoto and Hikosaka, 2007). Nigrostriatal and mesolimbic dopaminergic projections to the dorsolateral and ventral striatum, respectively, are used to modulate synaptic plasticity according to temporal difference learning.

and update its predictions such as to minimize these errors:

$$V(S_{t-1})^{\text{new}} = V(S_{t-1})^{\text{old}} + \eta\delta_t$$

where η is a learning rate ($0 < \eta < 1$; for a more detailed discussion of TD learning, see Barto, 1995; Niv and Schoenbaum, 2008). The elegance of this learning scheme is that local information (immediate rewards and stored estimations of values of the current and immediately preceding states) can be used to incrementally learn correct *long-run* value predictions.

Armed with long-run predictive values, the Critic can now coach the Actor. In fact, it turns out that the Actor can also learn from the same prediction error that the Critic computes in order to train its own values. This is because all that the Actor needs to know after every action, is whether this action improved its prospects for reward in the future, or not, that is, whether $r(S_t) + V(S_t) > V(S_{t-1})$, i.e., the prediction error δ_t is positive, or $r(S_t) + V(S_t) < V(S_{t-1})$ and the prediction error δ_t is negative, respectively. Actions that result in an increase in expected rewards should be repeated more often when in a similar situation, and vice versa for those which result in a decreased expectation of reward. Thus the same learning rule can be used to update the Actor's action propensities $\pi(a|S_t)$:

$$\pi(a|S_{t-1})^{\text{new}} = \pi(a|S_{t-1})^{\text{old}} + \eta\delta_t$$

where $\pi(a|S_{t-1})$ is the propensity to perform action a when in state S_{t-1}.

To recap, in the Actor/Critic architecture the Critic stores and learns state values and uses these to compute prediction errors, by which it updates its own state values. These same prediction errors are also conveyed to the Actor who stores and learns an action selection policy. The Actor uses the Critic's prediction error as a surrogate reinforcement signal with which it can improve its policy. In this particular division of labor, the "environment" sees only the output of the Actor, (i.e., the actions), however, rewards from the environment are only of interest to the Critic.

ACTOR/CRITIC IN THE BRAIN: THE BASAL GANGLIA

The Actor/Critic architecture is by no means the only solution to the credit assignment problem – it is certainly not the most efficient solution or computationally sound option (see Sutton and Barto, 1998 for a variety of other reinforcement learning algorithms). However, converging behavioral, anatomical and physiological evidence links the Actor/Critic architecture to habitual behavior of animals and humans, and to action selection mechanisms in the basal ganglia (Barto, 1995; Houk et al., 1995; Joel et al., 2002).

The main neural structures implicated in associative learning and action selection are the basal ganglia (in particular the striatum), limbic subcortical structures (the amygdala and the hippocampus), and prefrontal cortical areas. Generally speaking, these can be viewed as the decision-making interface between sensory (input) and motor (output) areas of the brain. The striatum, which is the input structure of the basal ganglia, receives convergent topographically organized inputs from motor, sensory, and limbic prefrontal cortical areas, as well as from the basolateral nuclei of the amygdala, the hippocampus, and the sensory thalamus. The basal ganglia then provide a positive feedback loop to the frontal cortex through cortico-striatal-pallido-thalamo-cortical loops (Albin et al., 1989; Alexander et al., 1986; Joel and Weiner, 1994; Parent and Hazrati, 1993). Thus the striatum is well positioned to identify meaningful associations between cues, responses and outcomes and to use this information to influence planning and action selection in the cortex (Joel and Weiner, 1999). Dopaminergic projections to the striatum are thought to play a critical role in this process by signaling reward prediction errors (Montague et al., 1996; Schultz et al., 1997) and modulating plasticity in cortico-striatal synapses (Wickens, 1990; Wickens et al., 2003).

Early mappings of the Actor/Critic model onto the basal ganglia identified the value-learning part of the Critic and the policy learning Actor with the "patch" (or striosome) and "matrix" (or matrisome) subpopulations of striatal medium spiny neurons, respectively (Brown et al., 1999; Contreras-Vidal and Schultz, 1999; Houk et al., 1995; Suri, 2002; Suri and Schultz, 1999; Suri et al., 2001), and the prediction error signal as carried by dopaminergic projections to the striatum (e.g., Barto, 1995; Houk et al., 1995). The plausibility of the patch/matrix mapping was later challenged on anatomical grounds (Joel et al., 2002), specifically in relation to how values in the Critic could affect the computation of a prediction error in dopaminergic nuclei, which is then conveyed to both Critic and Actor. Instead, current findings suggest a mapping in which the ventral subdivision of the striatum (including the nucleus accumbens) embodies the value learning part of the Critic that influences dopaminergic prediction errors, and the dorsal striatum is the Actor (Joel et al., 2002; but see Atallah et al., 2007 for a slightly different suggestion). While much evidence supports this suggestion, still more recent data advocate a refinement of the Actor to include only the dorsolateral striatum (with the dorsomedial striatum implicated in a different form of goal-directed action selection that cannot be supported computationally by an Actor/Critic architecture; Balleine, 2005; Yin et al., 2005). In the following we will briefly review the evidence supporting this ventral–dorsolateral striatal mapping of the Actor/Critic model, depicted with more detail in **Figure 1B**.

Recent work on the striatum has emphasized its division into different substructures based on connectivity patterns and functional roles (see **Figure 2**). The primary division is into dorsal and ventral parts (or dorsolateral and ventromedial; Voorn et al., 2004). The dorsal striatum is further comprised of two subparts: the caudate nucleus (or its homologue in rats, the dorsomedial striatum) and the putamen nucleus (or the dorsolateral striatum in rats). Similarly, within the ventral striatum the nucleus accumbens can be subdivided into core and shell compartments. This anatomical parcellation is in line with a previously suggested functional division of the basal ganglia into limbic (accumbal), associative (dorsomedial striatal) and sensorimotor (dorsolateral striatal) loops (Joel and Weiner, 1994; Parent and Hazrati, 1993). Dopaminergic projections from the ventral tegmental area (VTA; so-called "mesolimbic dopamine") target the ventral striatum (including the nucleus accumbens) while "nigrostriatal" dopamine arising from the substantia nigra pars compacta (SNc) targets the dorsal striatum (Amalric and Koob, 1993).

The ventral striatum, at the heart of the limbic corticostriatal loop, is well positioned to support learning of predictive values (as in the Critic). Afferents from so-called "limbic" areas, the basolateral amygdala and hippocampus, as well as from prefrontal cortical areas (Voorn et al., 2004), can convey information about the current context and stimuli, and their affective values. Efferents from the ventral striatum, in turn, target dopaminergic neurons, thus state values can be incorporated into the prediction error signal. Importantly, the ventral striatum influences activity in both dopaminergic nuclei, that is, it projects to those dopaminergic neurons that project back to the ventral striatum and to those that project to the more motor-related dorsal striatum (Haber et al., 1990, 2000; Joel and Weiner, 1999, 2000; Lynd-Balta and Haber, 1994;

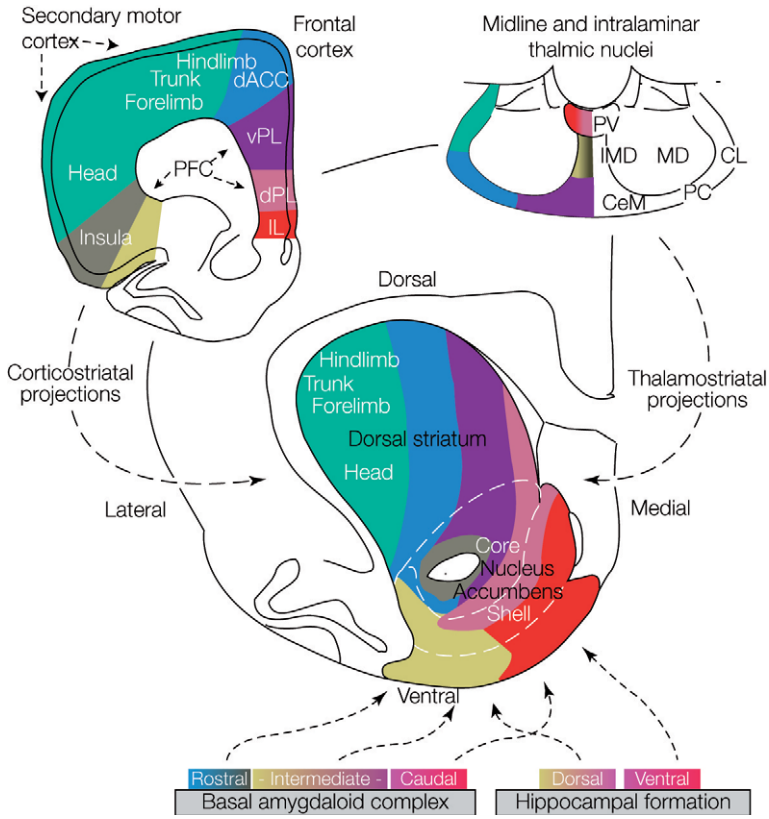

Figure 2 | **Anatomical organization of the rat striatum and its afferents.** In the center is a transverse section of the rat forebrain showing the striatum. Converging excitatory afferents from the frontal cortex (upper left), midline and intralaminar thalamic nuclei (upper right), basal amygdaloid complex (lower left) and hippocampal formation (lower right) project topographically to dorsomedial-to-ventrolateral zones. Frontal cortical areas and their corresponding striatal projection zones are shown in the same colors. Thalamic, amygdaloid and cortical afferents nicely align on one functional–anatomical organization. Abbreviations: ACC, anterior cingulate cortex; IL, infralimbic cortex; PFC, prefrontal cortex; dPL and vPL, dorsal and ventral prelimbic cortices; CeM, CL, IMD, MD, PC and PV, central medial, central lateral, intermediodorsal, mediodorsal, paracentral and paraventricular thalamic nuclei, respectively (adapted from Voorn et al., 2004).

Nauta et al., 1978). Due to this strategic positioning, the ventral striatum has been likened to a "limbic-motor interface" (Mogenson et al., 1980). In terms of the suggested Actor/Critic mapping, the ventral striatum is in a prime position to influence prediction error learning signals for both the Actor and the Critic, as is necessary in the model (see **Figure 1**).

In accord with these suggestive anatomical data, behavioral, pharmacological, and neural imaging results have implicated the ventral striatum in evaluation and prediction, and the dorsolateral striatum in action selection and execution (e.g., O'Doherty et al., 2004; Pessiglione et al., 2006; Robbins et al., 1989; **Figure 1B**). More specifically, the ventral striatum has been shown to be involved in reward anticipation, in attributing value to Pavlovian stimuli, and in mediating the ability of the affective value of anticipated outcomes to affect instrumental performance, as in paradigms such as Pavlovian-instrumental transfer and conditioned reinforcement (Balleine, 2005; Cardinal et al., 2002; Everitt et al., 1991; O'Doherty et al., 2003). A further division of labor between the nucleus accumbens core and shell, putatively involved in general value predictions (as in the Critic) and specific outcome values (as is necessary for correct goal-directed responding), respectively, is also in line with current data (Cardinal et al., 2002; Corbit et al., 2001; Hall et al., 2001). The dorsolateral striatum, on the other hand, has been implicated in habit learning and in the control of habitual responding (Atallah et al., 2007; Bailey and Mair, 2006; Balleine, 2005; Knowlton et al., 1996; Racht-Delatour and El Massioui, 2000; Yin et al., 2004, 2006), and is also the major site of dopamine loss in idiopathic Parkinson's disease, with its predominantly motor-related symptoms (e.g., Kish et al., 1988). Finally, recent results show that disconnecting the dopaminergic influence of the accumbens core on the dorsolateral striatum impairs the execution of habitual responding for a second order reinforcer predictive of cocaine (Belin and Everitt, 2008). This effect may be general and not limited to drug-seeking behavior, thus illustrating the importance of these spiraling projections from ventral to dorsal areas.

It has proven more difficult to garner direct support for this hypothesized functional subdivision from electrophysiological studies, as these have uncovered a bewildering wealth of representations in the striatum – essentially every possible task-related event can be shown to be represented by striatal medium spiny neurons (e.g., Schultz et al., 2000; Tremblay et al., 1998). Still, prominent in these are representations of outcome anticipation in ventral striatal areas (Schultz et al., 1992; Tremblay et al., 1998; Williams et al., 1993) and those of actions and of action "chunks" in the dorsal striatum (Kimura, 1986, 1995; Jog et al., 1999; Schultz and Romo, 1988; Ueda and Kimura, 2003).

The effects of dopamine on striatal neurons are also multifaceted. Structurally, glutamatergic cortical afferents and midbrain dopaminergic inputs converge onto common dendrites of striatal medium spiny neurons. Functionally, and important for the Actor/Critic model, dopamine has been shown to affect plasticity in cortico-striatal synapses, predominantly through D1 receptors, forming what has been termed a "three factor learning rule" in which the coincidence of presynaptic and postsynaptic activity causes LTP if a modulatory dopaminergic signal is also present, and LTD otherwise (Calabresi et al., 2000; Houk et al., 1995; Wickens, 1990; Wickens and Kotter, 1995; Wickens et al., 1996). For instance, in the accumbens core, co-infusion of low doses of a D1 receptor antagonist and an NMDA blocker impaired acquisition of instrumental learning (lever pressing for food),

even though these low doses were not sufficient to induce an impairment when infused alone (Smith-Roe and Kelley, 2000).

In sum, the evidence is consistent with a prediction and action selection system in which phasic dopamine signals a TD prediction error signal, which trains value predictions (as in the Critic) via projections from the VTA to ventral striatal and frontal target areas while simultaneously reinforcing the current action-selection policy (as in the Actor) through SNc projections to the dorsolateral striatum. Indeed, within a broader scheme, one might view the basal-ganglia/dopamine system as a system of multiple Actors and Critics operating in parallel, which can collaborate or compete to control behavior (Daw et al., 2005). In this, the dorsolateral striatum may support an Actor that relies on action propensities that are learned incrementally through trial and error (as described above), suggestive of inflexible stimulus-response habits that are independent of specific outcome predictions. By contrast, the dorsomedial striatum may support a second Actor that relies on computations of action values based on so-called response-outcome representations, thereby allowing goal-directed action selection that is more flexible and controlled by detailed outcome expectations (and not only scalar valued expectations in some common currency). Finally, Pavlovian responding may result from the operation of a third Actor or may be a direct consequence of value predictions (Dayan et al., 2006), thus placing the Pavlovian "Actor" in value learning areas such as the ventral striatum and the amygdala (Balleine and Killcross, 2006).

Whether the value learning Critic system is also divisible into a scalar prediction Critic (as necessary for the traditional Actor/Critic architecture), perhaps in the nucleus accumbens core and central nucleus of the amygdala, and a second Critic supporting more specific predictions of outcomes (as in goal-directed behavior), perhaps in the accumbens shell, basolateral amygdala and/or orbitofrontal cortex, is less clear[2]. Also unclear is the role of dopamine in training the non-habitual Actors, and indeed whether there are outcome-specific and outcome-general prediction errors. In any case, for our current purposes the rather well-substantiated mapping of the dorsolateral striatum to a habitual Actor, and the nucleus accumbens core to the scalar value learning Critic (that can also support preparatory Pavlovian responses) suffices. This is because the data of Takahashi et al. (2007), to which we will now turn, were from these two areas.

If indeed the dorsolateral and ventral striatum and their corresponding dopaminergic signals function in such a highly integrated manner as suggested by the Actor/Critic architecture (for challenges to this view, see Dayan and Balleine, 2002), then small perturbations could be expected to have pronounced effects on learning and ultimately on normal, adaptive (habitual) responding. In the next section, we briefly describe evidence from Takahashi et al. (2007) showing that exposure to an addictive drug causes massive changes in putative correlates of associative learning in this system. Thereafter we will consider these changes within an Actor/Critic architecture and use this model to suggest potential mechanisms whereby cocaine might be having its effects.

COCAINE SENSITIZATION: DIFFERENTIAL EFFECTS ON THE DORSAL/VENTRAL DIVIDE

To study the long-term effects of cocaine on signaling of predicted value in the striatum, Takahashi et al. (2007) recorded single-unit activity in the ventral and dorsolateral striatum of rats that had been sensitized to cocaine approximately 1–3 months earlier (**Figure 3A**). Recordings were made as these rats and saline-treated control animals learned and reversed a series of novel go, no-go odor discrimination problems (**Figure 4A**). In each problem, a pair of novel odor cues were paired with a sucrose reward and a quinine punishment (Schoenbaum et al., 1999b). Correct performance depended on acquiring associations between the odor cues and the rewarding and aversive outcomes, presumably resulting in the attribution of motivational value or significance to the cues. In addition, the rats were required to associate each cue with a different response (go or no-go) in order to solve the discrimination and each reversal (**Figure 4A**). Previous work has shown that rats sensitized to or trained to self-administer cocaine exhibit a long-lasting reversal learning impairment in this task (Calu et al., 2007; Schoenbaum et al., 2004). Consistent with this, cocaine-treated rats in the current study successfully reversed a much lower proportion of the odor problems to which they were exposed than the saline-treated controls.

In the context of this behavior, Takahashi et al. (2007) found that in control rats neurons in both ventral and dorsolateral striatum fired to the cues based on their associations with subsequent responses and/or outcomes. This cue-selective activity developed with learning and reversed in both regions when the meaning of the cues was switched

[2]The evidence most in favor of such a subdivision of the Critic is rather circumstantial, deriving from the existence of two distinct forms of Pavlovian behavior (Balleine and Killcross, 2006; Dickinson and Balleine, 2002) – preparatory behavior that is motivationally specific but otherwise rather outcome-general (as in approach and withdrawal), and consummatory responses that are specific to the predicted stimulus (such as licking, salivating, etc.).

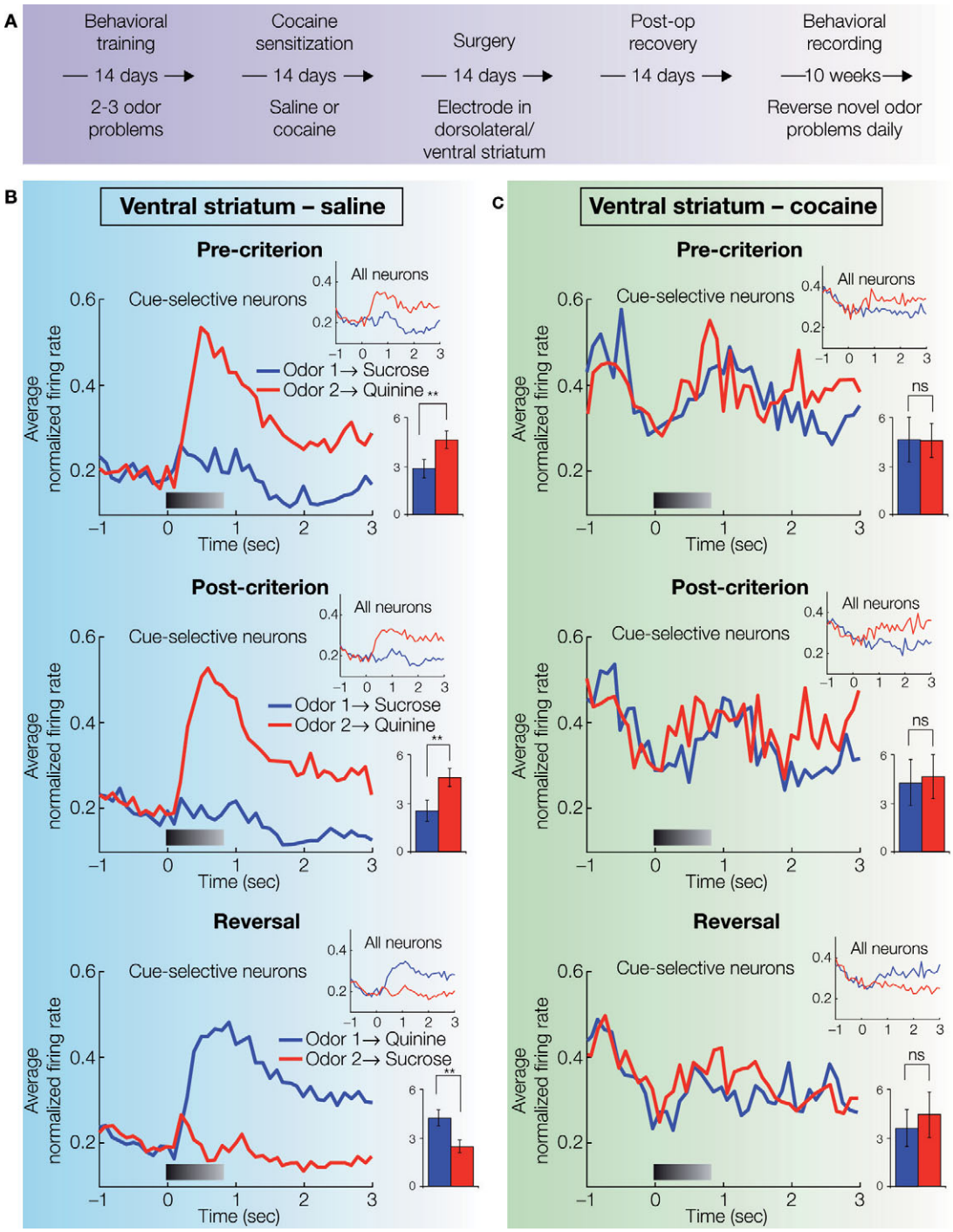

Figure 3 | **Single-unit activity in ventral striatum in saline- and cocaine-treated rats during discrimination and reversal learning.** ***(A)*** Experimental timeline showing the relative timing of behavioral training, cocaine sensitization and recording. ***(B,C)*** Average activity in populations of neurons recorded in ventral striatum (nucleus accumbens core) during acquisition and reversal of a series of odor discrimination problems (only successful reversals were included), during pre-criterion trials, post-criterion trials and post-reversal. Activity was normalized to the maximum firing of each neuron during odor sampling. The left column ***(B)*** shows activity in saline-treated control rats and the right column ***(C)*** shows activity in cocaine-treated rats. In each case, population responses are shown separately for all neurons (small insets) and for cue-selective neurons (large plots; cue-selective neurons were defined as any neuron with differential activity during odor sampling after learning at $p < 0.05$, ANOVA). In blue is activity on trials with odor 1 (which predicted sucrose before reversal), and in red is activity on trials with odor 2 (which predicted quinine before reversal). Gray shading in each histogram indicates the approximate timing of odor sampling. Bar graphs show averaged firing rates (spikes per second) for each odor in each phase in the cue-selective populations (*, significant difference at $p < 0.05$; **, significant difference at $p < 0.01$ or better, ANOVA). The neural population in saline-treated controls showed strong firing to the odor cue that predicted the aversive quinine outcome in pre-criterion trials, post-criterion trials, and after reversal. In contrast, the neural population in cocaine-treated rats was not cue-selective in either phase (adapted from Takahashi et al., 2007).

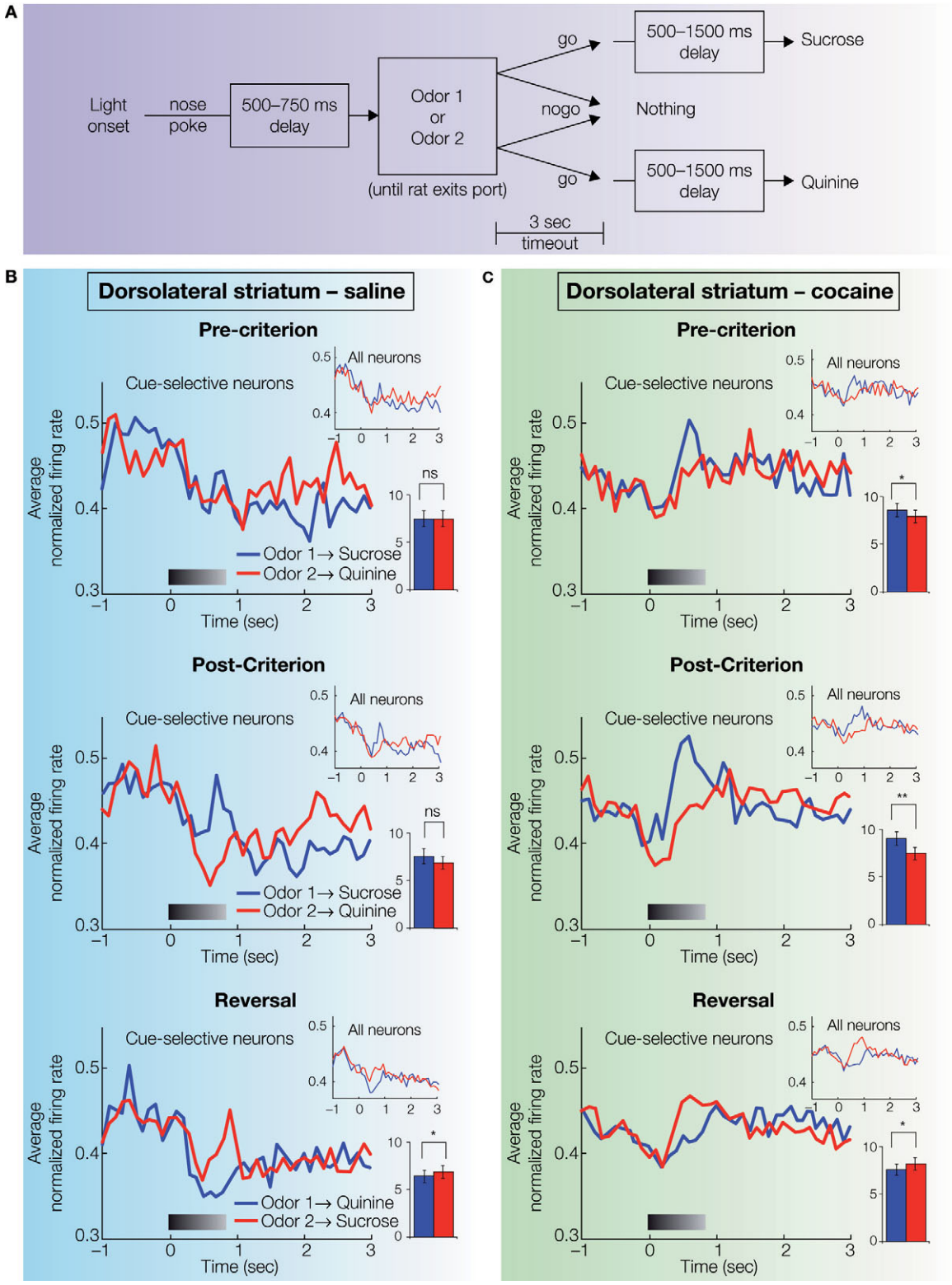

Figure 4 | **Single-unit activity in dorsolateral striatum in saline- and cocaine-treated rats during discrimination and reversal learning.** ***(A)*** Timing of events within each trial. After reversal, the mapping between odor 1 and odor 2 and the sucrose and quinine outcomes was reversed. Odor identities were counterbalanced across rats. ***(B,C)*** Average activity in populations of neurons recorded in dorsolateral striatum during acquisition and reversal of a series of odor discrimination problems (conventions as in Figure 3). The neural population in cocaine-treated group exhibited stronger phasic firing to the odor cue that predicted the sucrose outcome than the population in saline-treated controls. This difference was especially evident during pre-criterion and post-criterion trials (adapted from Takahashi et al., 2007).

(**Figures 3B and 4B**). However several key differences emerged between the two areas, consistent with their putative roles in the Actor/Critic models. Cue-selectivity in ventral striatum developed rapidly during learning, preceding the development of accurate responding, and firing was much higher to the odor cue that predicted the more motivationally-significant aversive quinine outcome (**Figure 3B**). In contrast, cue-selectivity in dorsolateral striatum developed later, only alongside accurate differential responding (**Figure 4B**). These differences fit well with the proposal that ventral striatum learns the value of the cue, so that it is then able to coach the dorsolateral striatum regarding the proper action to execute in response to each odor cue.

Consistent with the proposal that addiction might involve changes in striatal function, prior cocaine-treatment had a significant impact on the neural correlates in both striatal regions (**Figures 3C and 4C**, compare to **Figures 3B and 4B**). First, after learning, cocaine-treated rats had substantially fewer cue-selective neurons in ventral striatum than saline controls. The absence of such neurons in cocaine-treated rats was evident in the population responses (**Figure 3C**), which did not exhibit cue-selectivity in any phase of training. By contrast, cocaine-treatment caused a small but significant increase in the number of cue-responsive neurons in dorsolateral striatum, especially to the positive odor cue that drove most of the neural activity observed in controls. The effect is somewhat evident in the population responses in **Figure 4C**, where the relatively weak and slow to develop response to the positive odor cue seen in controls is more robust in the cocaine-treated rats. The stronger response to the positive odor in cocaine-treated rats was particularly evident in the pre-criterion trials, thus differential activity appeared earlier in cocaine-treated rats in dorsolateral striatum.

These results show that cocaine sensitization shifted the balance of encoding between these two regions, abolishing the strong cue-selectivity normally present in ventral striatum while marginally enhancing the relatively weak cue-selectivity normally present in dorsolateral striatum. In the next section, we will consider how such an effect might be interpreted within the Actor/Critic model of striatal function.

A COACH GONE AWRY AND AN ACTOR RUNNING LOOSE

According to the above mapping between an Actor/Critic architecture and the subparts of the striatum, one consequence of radically reduced cue selectivity in the ventral striatum may be the severe degradation (or even elimination) of the feedback from the Critic to the Actor. In other words, if cue-evoked activity in ventral striatum is in fact signaling information about state values, as proposed for the Critic, then Takahashi et al.'s (2007) results show that cocaine sensitization disrupts this function. Within an Actor/Critic model, this would have profound effects on the ability of the Actor to learn to select the appropriate actions. In the following, we discuss two different ways that such a disturbance might be manifest, and their implications for action selection. In the Discussion we further develop ideas for future experiments targeted at testing these two alternatives more directly.

How can we envision an Actor/Critic network without a functioning Critic? The simplest possibility, depicted in **Figure 5A**, is that the Critic fails to learn or to represent state values properly, that is, the $V(S_t)$ signals that it produces are meaningless. These would then be integrated into the prediction error that trains the Actor, and would lead to a perturbed learning process. This scenario is generally consistent with the results of Takahashi et al. (2007) in that cue-evoked activity in the ventral striatum of cocaine-sensitize rats did not distinguish between the two cues. We can thus model this as a situation in which all $V(S_t)$ are the same, regardless of the state S_t. In this case, the prediction error $\delta_t = r(S_t) + V(S_t) - V(S_{t-1})$ reduces to $\delta_t = r(S_t)$, i.e., the prediction error is equal to the immediately available reward only.

Although it is obviously naïve to envision cocaine sensitization as completely wiping out any representation of $V(S_t)$, we think it is useful to consider such an extreme situation in order to understand the potential consequences of a deficit in the Critic's representations. Specifically, we'd like to ask what are the consequences of this hypothesized deficient training signal for the Actor, i.e., for representations in the dorsolateral striatum and for overt behavior? On the one hand, in tasks with outcomes that depend on a series of actions, that is, those scenarios in which the credit assignment problem is severe and an intact Actor/Critic architecture is necessary for proper learning, a learning signal that is based at every time-point only on immediate reward will not be sufficient to learn the task. This predicts that cocaine sensitization would have a large detrimental effect on subsequent habitual learning and action selection in such tasks. On the other hand, in those (simpler to learn) tasks in which actions are immediately rewarded in an unambiguous way, learning may actually be faster. That is because it will not be hindered by incorrect estimations of future values which are common early on in the learning process.

Odor discrimination learning is of the latter kind – odors predict the immediate outcome (sucrose

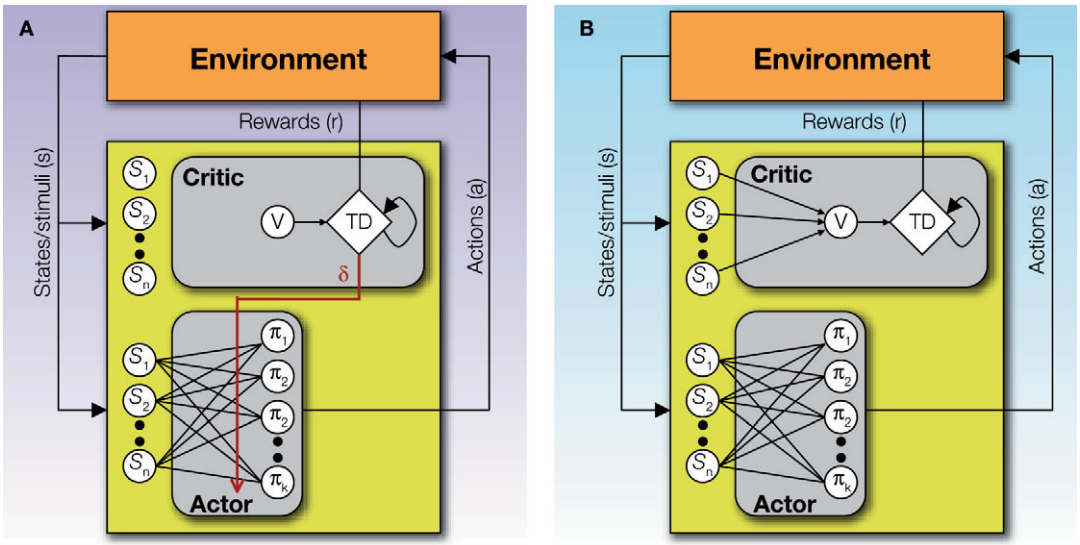

Figure 5 | **Two possible realizations of an Actor/Critic network with a dysfunctional Critic.** ***(A)*** The Critic is unable to learn or represent a meaningful mapping between states and their values (depicted is the extreme case of similar values for all states), thus the prediction error signal δ that is used to train the Actor comprises only of the current reward. ***(B)*** A deficient prediction error signal disrupts learning in both the Critic and the Actor (depicted is the extreme case of no prediction error whatsoever).

or quinine) to a go or a no-go response. In this sense, it is perhaps not surprising that performance was not disrupted after cocaine sensitization, and that representations in the dorsolateral striatum were intact, or even slightly enhanced. Indeed in a prior study in which initial learning was carefully examined, cocaine-treated rats showed a small but nearly significant improvement in acquisition of the initial discrimination problem learned (Schoenbaum et al., 2004), and we have observed a similar facilitation in accumbens-lesioned rats (Schoenbaum and Setlow, 2003). While such facilitations are difficult to observe in recording work, where rats have developed a "learning set" for these problems, these prior observations are consistent with the idea that removal of the part of the Critic that is located in ventral striatum may actually enhance learning of simple discriminations.

Furthermore, that cocaine-sensitized rats reversed fewer of the discriminations successfully may, in fact, hint at stronger learning of the initial discrimination, which then impeded later reversal of the behavioral policy. Indeed, learning with a persistent reward signal (i.e., one that does not come to be "predicted away" by predictive cues) could lead to stronger habitual learning, to the point of maximal action propensities. This idea would be consistent with other evidence showing that reversal deficits after cocaine are associated with inflexible encoding in the basolateral amygdala (Stalnaker et al., 2007).

A second option, depicted in **Figure 5B**, is that cocaine sensitization affects the prediction error signal itself. An invalid error signal (or, considering once again the extreme case – no prediction error signal) would disrupt learning in both the Critic and the Actor and thus would account for the loss of cue-selectivity in the ventral striatum. However, this idea is not in line with the finding that both behavior and the representations in the dorsolateral striatum were intact, unless there is some other means by which action propensities and dorsolateral striatal representation can be learned, at least in cases in which rewards are immediate. This latter suggestion is not entirely speculative – as mentioned, recent work has highlighted the existence of multiple learning and action selection systems in the brain, presumably all working in parallel and interacting (whether in competition or in synergy) to control behavior (Balleine 2005; Daw et al., 2005; Johnson et al., 2007). Pavlovian "approach" and "withdrawal" responses, perhaps acquired through prediction learning in the central and basolateral nuclei of the amygdala, and/or in the nucleus accumbens shell (from which Takahashi et al., 2007 did not record), could suffice to generate the correct behavioral response (Dayan et al., 2006), and could perhaps provide surrogate training signals for the Actor. Such alternative training could conceivably result in representations that are more resistant to reversal, due to the absence of a flexibly adapting dopaminergic training signal.

Of the two options we suggest, it is not immediately clear which is more in line with the effects of cocaine sensitization. This is partly due to the requirements of the task used by Takahashi et al. (2007), a go/no-go odor discrimination task, with

immediate outcomes in each trial. This task was chosen because rats can readily learn it (and perform reversals), which allows for clear assessment of behavioral and neural implications of different treatments on the learning process itself. Indeed, this task has been widely used to record in other areas, and robust behavioral and neural effects have been shown with previous pharmacological or lesion manipulations (Gallagher et al., 1999; Saddoris et al., 2005; Schoenbaum et al., 1998, 1999a, 2003; Setlow et al., 2003; Stalnaker et al., 2006, 2007). However, that this task is learned with such ease is perhaps because of its simple structure: only one action is required to obtain (or avoid) an outcome (see **Figure 4A**), rather than a series of actions. Moreover, this action is not opposed to the Pavlovian tendency to approach predictors of reward and avoid predictors of punishment. This is likely also the reason that cocaine-sensitized rats with severely disrupted ventral striatal representations were still able to learn new discriminations and reversals, and basically display intact behavior in this task. Indeed, it seems that in tasks that provide a more difficult test-bed for the Actor/Critic mechanism, both deficits should show more severe implications for learning and action selection, much beyond those seen in Takahashi et al.'s (2007) results.

DISCUSSION

Takahashi et al.'s (2007) study showed that cocaine sensitization disrupts encoding in the ventral striatum, while leaving dorsolateral striatal encoding and behavior mostly intact. Based on the idea that the dorsolateral and ventral striatum may function as an Actor and a Critic, respectively, we have suggested two ways in which these results can be interpreted within an Actor/Critic framework. Although this is certainly not the only framework within which to consider these data, the Actor/Critic framework has strong empirical support and also makes very specific predictions about the underlying mechanism. In particular, the two ways that we suggest these changes might arise differ in whether the primary (and thus treatable) cause lies within ventral striatum itself or is in the incoming dopaminergic projections.

Several straightforward directions of future research thus come to mind – extensions which would allow us to tease apart the two possible disruptions. First, if ventral striatal representations are indeed used to generate a training signal for the dorsolateral striatum which takes into account delayed as well as immediate rewards, a task in which outcomes are dependent on a sequence of actions should show grossly disrupted learning and performance after cocaine sensitization. At the very least, even if behavioral impairments are masked by other (intact) action selection systems in the brain, we expect to see disrupted representations in the dorsolateral striatum in this case. Second, recordings from dopaminergic neurons should reveal an altered prediction error signal after cocaine sensitization: the signal should either be dominated by immediate rewards (according to the option in **Figure 5A**) or be altogether uninformative or absent (as in **Figure 5B**). Third, if the prediction error signal is indeed wholly disrupted as in the second option above, a task in which Pavlovian conditioning is not sufficient to produce the correct behavior could reveal additional behavioral deficits. Finally, in this case (as well as in general), it would be extremely interesting to test the implications of cocaine sensitization on representations in the shell of the nucleus accumbens and the dorsomedial striatum.

As we have already noted, cocaine-induced impairments in reversal tasks or in tasks that model aspects of reversal learning appear to reflect not an absence of learning but rather abnormally strong initial learning. That is, addicts and animals exposed to cocaine show normal responding before reversal but then fail to change behavior after reversal. In the present study, these reversal deficits seemed to be linked to enhanced signaling in the dorsolateral striatum. However, we have previously demonstrated that they are also associated with persistent miscoding of the original associations in the basolateral amygdala (Stalnaker et al., 2007). Moreover, removal of the basolateral amygdala is sufficient to restore normal performance (Stalnaker et al., 2007), suggesting that this aberrant encoding is the proximal cause of the cocaine-dependent reversal deficit. These results are also consistent with the observation that cue-evoked relapse is ameliorated by infusions of agents into the amygdala that disrupt memory reconsolidation during exposure to drug cues (Lee et al., 2005).

Thus, drug exposure does not seem to prevent learning in general, as would be expected if error signals were eliminated. Instead, it seems to make learning more difficult when delayed or probabilistic outcomes must be tallied. For instance, addicts and drug-treated animals show increased sensitivity to delay of reward, behaving more "impulsively", and also show increased sensitivity to reward magnitude (Coffey et al., 2003; Roesch et al., 2007; Simon et al., 2007). This pattern of results suggests interference with the Critic, but supports the idea that at least some part of the prediction error signal is intact, and continues to support learning. In fact, reliance on a prediction error that is dominated by immediate rewards would indeed result in increased impulsivity and sensitivity to reward

magnitude. However, these are also consistent with increasing reliance on a Pavlovian learning system, which could be an obvious consequence of any type of disruption of the instrumental action-selection mechanism.

Interestingly, although it is not well-modeled with current theories of instrumental choice, cocaine also appears to cause significant changes in how information about expected outcomes is signaled by frontal areas. We have previously demonstrated that cocaine-treated rats are unable to use information about outcomes to guide behavior in a Pavlovian setting (Schoenbaum and Setlow, 2005), and Killcross and colleague have shown analogous effects of amphetamine on outcome-guided behavior in an instrumental setting (Nelson and Killcross, 2006). Similar deficits in the control of Pavlovian and instrumental behavior are caused by damage to the orbitofrontal and medial prefrontal areas respectively (Corbit and Balleine, 2003; Gallagher et al., 1999; Killcross and Coutureau, 2003), and are thought to reflect the role of these areas in signaling information about expected outcomes, potentially to both instrumental and Pavlovian systems. Thus drug exposure may also disrupt outcome signaling by these critical prefrontal areas. In accord with this hypothesis, we have found that orbitofrontal neurons in cocaine-treated rats fail to signal expected outcomes during cue-sampling (Stalnaker et al., 2006).

This proliferation of drug-induced deficits suggests yet a third possible explanation to the results we have described above: the loss of differential encoding of cues in the ventral striatum could be the result of disrupted representations carried by its cortical afferents. Thus, a common cause might be at the heart of the orbitofrontal, amygdalar, and ventral striatal deficits, or one of these may be influencing the others, to their detriment. However, the preservation or even enhancement of the dorsolateral striatal representations argues against this idea, especially if the core of the deficit were in higher cortices that presumably modulate not only the limbic corticostriatal loop, but also the associative and motor loops (Joel and Weiner, 1999). That the dorsolateral striatal representations remained intact thus suggests that the problem may not be as severe as an insidious lack of representation in the prefrontal cortex. Still, the relationships and interactions between the different drug-induced deficits remain as an intriguing question for future research.

In sum, the pervasive and detrimental effects of (even mild forms of) drug abuse are astounding. We have suggested here that combining experimental findings with a computational analysis can allow not only for a system-level interpretation of electrophysiological results and their relationship to behavioral deficits, but can also suggest a wealth of new experiments. Indeed, it sometimes seems that the questions outnumber the answers. By bringing theoretical models to bear on concrete findings, we can only hope to make another small step in the direction of understanding (and ultimately curing) drug addiction.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to Daphna Joel and Peter Dayan for many helpful comments on an earlier draft, and to the reviewers for their constructive comments on an earlier version of this paper. This work was funded by Human Frontiers Science Program Fellowships to Yael Niv and to Yuji Takahashi, and by an R01-DA015718 grant to Geoffrey Schoenbaum from the National Institute on Drug Abuse (NIDA).

REFERENCES

Albin, R. L., Young, A. B., and Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends Neurosci.* 12, 366–375.

Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel organization of functional segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* 9, 357–381.

Amalric, M., and Koob, G. F. (1993). Functionally selective neurochemical afferents and efferents of the mesocorticolimbic and nigrostriatal dopamine system. *Prog. Brain Res.* 99, 209–226.

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

Bailey, K. R., and Mair, R. G. (2006). The role of striatum in initiation and execution of learned action sequences in rats. *J. Neurosci.* 26, 1016–1025.

Balleine, B. W. (2005). Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiol. Behav.* 86, 717–730.

Balleine, B. W., and Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.* 29, 272–279.

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In Models of Information Processing in the Basal Ganglia, J. C. Houk, J. L. Davis and D. G. Beiser, eds (Cambridge, MA, MIT Press), pp. 215–232.

Barto, A. G., Sutton, R. S., and Anderson, C. W. (1983). Neuron-like adaptive elements that can solve difficult learning control problems. *IEEE Trans. Syst. Man Cybern.* 13, 834–846.

Barto, A. G., Sutton, R. S., and Watkins, C. J. C. H. (1989). Sequential decision problems and neural networks. In Advances in Neural Information Processing Systems, Vol. 2, D. S. Touretzky, ed. (Cambridge, MA, MIT Press), pp. 686–693.

Bechara, A., and Damasio, H. (2002). Decision-making and addiction (part I): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia* 40, 1675–1689.

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

Brown, J., Bullock, D., and Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J. Neurosci.* 19, 10502–10511.

Calabresi, P., Gubellini, P., Centonze, D., Picconi, B., Bernardi, G., Chergui, K., Svenningsson, P., Fienberg, A. A., and Greengard, P. (2000). Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. *J. Neurosci.* 20, 8443–8451.

Calu, D. J., Stalnaker, T. A., Franz, T. M., Singh, T., Shaham, Y., and Schoenbaum, G. (2007). Withdrawal from cocaine self-administration produces long-lasting deficits in orbitofrontal-dependent reversal learning in rats. *Learn. Mem.* 14, 325–328.

Cardinal, R. N., Parkinson, J. A., Hall, G., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* 26, 321–352.

Christoph, G. R., Leonzio, R. J., and Wilcox, K. S. (1986). Stimulation of the lateral habenula inhibits dopamine containing neurons in the substantia nigra and ventral tegmental area of the rat. *J. Neurosci.* 6, 613–619.

Coffey, S. F., Gudleski, G. D., Saladin, M. E., and Brady, K. T. (2003). Impulsivity and rapid discounting of delayed hypothetical rewards in cocaine-dependent individuals. *Exp. Clin. Psychopharmacol.* 11, 18–25.

Contreras-Vidal, J. L., and Schultz, W. (1999). A predictive reinforcement model of dopamine neurons for learning approach behavior. *J. Comput. Neurosci.* 6, 191–214.

Corbit, L. H., and Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behav. Brain Res.* 146, 145–157.

Corbit, L. H., Muir, J. L., and Balleine, B. W. (2001). The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. *J. Neurosci.* 21, 3251–3260.

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

Dayan, P., and Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron* 36, 285–298.

Dayan, P., Niv, Y., Seymour, P., and Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Netw.* 19, 1153–1160.

Dickinson, T., and Balleine, B. W. (2002). The role of learning in the operation of motivational systems. In Learning, Motivation and Emotion, C. R. Gallistel, ed. (New York, NY, John-Wiley and Sons), pp. 497–533.

Everitt, B. J., Morris, K. A., O'Brien, A., and Robbins, T. W. (1991). The basolateral amygdala-ventral striatal system and conditioned place preference: further evidence of limbic–striatal interactions underlying reward-related processes. *Neuroscience* 42, 1–18.

Gallagher, M., McMahan, R. W., and Schoenbaum, G. (1999). Orbitofrontal cortex and representation of incentive value in associative learning. *J. Neurosci.* 19, 6610–6614.

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

Haber, S. N., Lynd-Balta, E., Klein, C., and Groenewegen, H. J. (1990). Topographic organization of the ventral striatal efferent projections in the rhesus monkey: an anterograde tracing study. *J. Comp. Neurol.* 293, 282–298.

Hall, J., Parkinson, J. A., Connor, T. M., Dickinson, A., and Everitt, B. J. (2001). Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behavior. *Eur. J. Neurosci.* 13, 1984–1992.

Houk, J. C., Adams, J. L., and Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In Models of Information Processing in the Basal Ganglia, J.C. Houk, J. L. Davis and D. G. Beiser, eds (Cambridge, MA, MIT Press), pp. 249–270.

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

Ji, H., and Shepard, P. D. (2007). Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABA(A) receptor-mediated mechanism. *J. Neurosci.* 27, 6923–6930.

Joel, D., Niv, Y., and Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw.* 15, 535–547.

Joel, D., and Weiner, I. (1994). The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. *Neuroscience* 63, 363–379.

Joel, D., and Weiner, I. (1999). Striatal contention scheduling and the split circuit scheme of basal ganglia-thalamocortical circuitry: from anatomy to behaviour. In Conceptual Advances in Brain Research: Brain Dynamics and the Striatal Complex, R. Miller and J. R. Wickens, eds (Newark, Harwood Academic Publishers), pp. 209–236.

Joel, D., and Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience* 96, 451–474.

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., and Graybiel, A. M. (1999). Building neural representations of habits. *Science* 286, 1745–1749.

Johnson, A., van der Meer, M. A. A., and Redish, A. D. (2007). Integrating hippocampus and striatum in decision-making. *Curr. Opin. Neurobiol.* 17, 692–697.

Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* 13, 400–408.

Kimura, M. (1986). The role of primate putamen neurons in the association of sensory stimuli with movement. *Neurosci. Res.* 3, 436–443.

Kimura, M. (1995). Role of basal ganglia in behavioral learning. *Neurosci. Res.* 22, 353–358.

Kish, S. J., Shannak, K., and Hornvkiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease: pathophysiologic and clinical implications. *N. Engl. J. Med.* 318, 876–880.

Knowlton, B. J., Mangels, J. A., and Squire, L. (1996). A neostriatal habit learning system in humans. *Science* 273, 1399–1402.

Kobayashi, Y., and Okada, K. (2007). Reward prediction error computation in the pedunculopontine tegmental nucleus neurons. *Ann. N. Y. Acad. Sci.* 1104, 310–323.

Lee, J. L., Di Ciano, P., Thomas, K. L., and Everitt, B. J. (2005). Disrupting reconsolidation of drug memories reduces cocaine-seeking behavior. *Neuron* 47, 795–801.

Lynd-Balta, E., and Haber, S. N. (1994). Primate striatonigral projections: a comparison of the sensorimotor-related striatum and the ventral striatum. *J. Comp. Neurol.* 345, 562–578.

Matsumoto, M., and Hikosaka, K. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447, 1111–1115.

Mogenson, G. J., Jones, D. L., and Yim, C. Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Prog. Neurobiol.* 14, 69–97.

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

Nauta, W. J. H., Smith, G. P., Faull, R. L. M., and Domesick, V. B. (1978). Efferent connections and nigral afferents of the nucleus accumbens septi in the rat. *Neuroscience* 3, 385–401.

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

Niv, Y., and Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends Cogn. Sci.* 12, 265–272.

O'Doherty, J., Dayan, P., Friston, K. J., Critchley, H., and Dolan, R. J. (2003). Temporal difference learning model accounts for responses in human ventral striatum and orbitofrontal cortex during Pavlovian appetitive learning. *Neuron* 38, 329–337.

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

Parent, A., and Hazrati, L. N. (1993). Anatomical aspects of information processing in primate basal ganglia. *Trends Neurosci.* 16, 111–116.

Pessiglione, M., Seymour, P., Flandin, G., Dolan, R. J., and Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442, 1042–1045.

Racht-Delatour, B. V. G., and El Massioui, N. (2000). Alleviation of overtraining reversal effect by transient inactivation of the dorsal striatum. *Eur. J. Neurosci.* 12, 3343–3350.

Robbins, T. W., Cador, M., Taylor, J. R., and Everitt, B. J. (1989). Limbic-striatal interactions in reward-related processes. *Neurosci. Biobehav. Rev.* 13, 155–162.

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

Saddoris, M. P., Gallagher, M., and Schoenbaum, G. (2005). Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. *Neuron* 46, 321–331.

Schoenbaum, G., Chiba, A. A., and Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.* 1, 155–159.

Schoenbaum, G., Chiba, A. A., and Gallagher, M. (1999a) Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J. Neurosci.* 19, 1876–1884.

Schoenbaum, G., Chiba, A. A., and Gallagher, M. (1999b) Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J. Neurosci.* 19, 1876–1884.

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

Schoenbaum, G., and Setlow, B. (2003). Lesions of nucleus accumbens disrupt learning about aversive outcomes. *J. Neurosci.* 23, 9833–9841.

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

Schoenbaum, G., Setlow, B., Saddoris, M. P., and Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* 39, 855–867.

Schultz, W., Apicella, P., Scarnati, E., and Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neurosci.* 12, 4595–4610.

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

Schultz, W., and Romo, R. (1988). Neuronal activity in the monkey striatum during the initiation of movements. *Exp. Brain Res.* 71, 431–436.

Schultz, W., Tremblay, L., and Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb. Cortex* 10, 272–283.

Setlow, B., Schoenbaum, G., and Gallagher, M. (2003). Neural encoding in ventral striatum during olfactory discrimination learning. *Neuron* 38, 625–636.

Simon, N. W., Mendez, I. A., and Setlow, B. (2007). Cocaine exposure causes long-term increases in impulsive choice. *Behav. Neurosci.* 121, 543–549.

Smith-Roe, S. L., and Kelley, A. E. (2000). Coincident activation of NMDA and dopamine D1 receptors within the nucleus accumbens core is required for appetitive instrumental learning. *J. Neurosci.* 20, 7737–7742.

Stalnaker, T. A., Roesch, M. R., Franz, T. M., Burke, K. A., and Schoenbaum, G. (2006). Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making. *Eur. J. Neurosci.* 24, 2643–2653.

Stalnaker, T. A., Roesch, M. R., Franz, T. M., Calu, D. J., Singh, T., and Schoenbaum, G. (2007). Cocaine-induced decision-making deficits are mediated by miscoding in basolateral amygdala. *Nat. Neurosci.* 10, 949–951.

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. *Neural Netw.* 15, 523–533.

Suri, R. E., Bargas, J., and Arbib, M. A. (2001). Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience* 103, 65–85.

Suri, R. E., and Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* 91, 871–890.

Sutton, R. S. (1988). Learning to predict by the method of temporal difference. *Mach. Learn.* 3, 9–44.

Sutton, R. S., and Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In Learning and Computational Neuroscience: Foundations of Adaptive Networks, M. Gabriel and J. Moore, eds (Boston, MIT Press), pp. 497–537.

Sutton, R. S., and Barto, A. G. (1998). Reinforcement Learning: An introduction. Cambridge, MA, MIT Press.

Takahashi, Y., Roesch, M. R., Stalnaker, T. A., and Schoenbaum, G. (2007). Cocaine exposure shifts the balance of associative encoding from ventral to dorsolateral striatum. *Front. Integr. Neurosci.* 1, 11. doi: 10.3389/neuro.07/011.2007.

Tremblay, L., Hollerman, J. R., and Schultz, W. (1998). Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J. Neurophysiol.* 80, 964–977.

Ueda, Y., and Kimura, M. (2003). Encoding of direction and combination of movements by primate putamen neurons. *Eur. J. Neurosci.* 18, 980–994.

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

Voorn, P., Vanderschuren, L. J. M. J., Groenewegen, H. J., Robbins, T. W., and Pennartz, C. M. A. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci.* 27, 468–474.

Wickens, J., and Kotter, R. (1995). Cellular models of reinforcement. In Models of Information Processing in the Basal Ganglia, J. C. Houk, J. L. Davis and D. G. Beiser, eds (Cambridge, MA, MIT Press), pp. 187–214.

Wickens, J. R. (1990). Striatal dopamine in motor activation and reward-mediated learning: steps towards a unifying model. *J. Neural Transm.* 80, 9–31.

Wickens, J. R., Begg, A. J., and Arbuthnott, G. W. (1996). Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience* 70, 1–5.

Wickens, J. R., Reynolds, J. N. J., and Hyland, B. I. (2003). Neural mechanisms of reward-related motor learning. *Curr. Opin. Neurobiol.* 13, 685–690.

Williams, G. V., Rolls, E. T., Leonard, C. M., and Stern, C. (1993). Neuronal responses in the ventral striatum of the behaving macaque. *Behav. Brain Res.* 55, 243–252.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.* 22, 505–512.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166, 189–196.

Received: 19 May 2008; accepted: 26 June 2008.

Citation: Front. Neurosci. (2008) 2, 1: 86–99, doi: 10.3389/neuro.01.014.2008