



Published in final edited form as:

Neuron. 2015 October 21; 88(2): 247–263. doi:10.1016/j.neuron.2015.08.037.

Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry

Ronald Keiflin¹ and Patricia H. Janak^{1,2}

¹Department of Psychological and Brain Sciences, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore MD 21218

²Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore MD 21205

Summary

Midbrain dopamine (DA) neurons are proposed to signal reward prediction error (RPE), a fundamental parameter in associative learning models. This RPE hypothesis provides a compelling theoretical framework for understanding DA function in reward learning and addiction. New studies support a causal role for DA-mediated RPE activity in promoting learning about natural reward; however, this question has not been explicitly tested in the context of drug addiction. In this review, we integrate theoretical models with experimental findings on the activity of DA systems, and on the causal role of specific neuronal projections and cell types, to provide a circuit-based framework for probing DA-RPE function in addiction. By examining error-encoding DA neurons in the neural network in which they are embedded, hypotheses regarding circuit-level adaptations that possibly contribute to pathological error-signaling and addiction can be formulated and tested.

Introduction

Drugs of abuse enhance dopamine (DA) function by directly or indirectly acting upon midbrain DA neurons to transiently increase extracellular concentrations of DA (Di Chiara and Imperato, 1988; Nestler, 2005; Sulzer, 2011). This DA enhancement has been considered to mediate multiple effects of addictive drugs, such as behavioral and motor activation, and, most prominently, the rewarding effects of these drugs. “Rewarding effects” may mean many things therefore more specific definitions are required (Berridge and Robinson, 2003). Indeed, in addition to producing positive hedonic reactions, rewards (both natural and pharmacological) engage motivational systems involved in the initiation and invigoration of action, as well as learning systems responsible for the reinforcement of the events (cues and actions) leading to the receipt of the rewarding outcome. Contemporary ideas of how DA is related to a drug’s rewarding effect emphasize DA’s role in motivation

Correspondence: Patricia H. Janak, Ph.D., patricia.janak@jhu.edu.

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

and/or reward learning (Berke and Hyman, 2000; Berridge, 2012; Berridge and Robinson, 1998; Bromberg-Martin et al., 2010; Everitt and Robbins, 2005; Salamone and Correa, 2012; Schultz, 2007 & 2013; Wise, 2004).

In this review, we explore the implications for addiction for one prominent idea of DA function, namely the signaling of reward prediction errors (RPEs) by DA neurons (DA-RPE). According to this view, DA neurons encode the discrepancy between reward predictions and information about the actual reward received (the RPE) and broadcast this signal to downstream brain regions involved in reward learning. Because drugs of abuse alter DA signaling, this theory has implications for addiction. In a seminal paper, Redish (2004) presented a computational delineation of how RPEs may play a critical role in addiction that has considerably shaped thinking in this area. In the ensuing years, theoretical models have been updated and new information regarding the neurobiology both of RPEs and of addiction has accrued, and the time is ripe for a re-examination of DA-RPE and addiction.

We start from the position that neural computation of an RPE by DA neurons contributes to learning about predictive cues and actions that lead to that reward, and ask, if this is true for normal learning, what role might DA-mediated RPEs play when subjects learn about drug-related cues and actions? Is there something different about RPE-mediated learning when the reward is an addictive drug and how does that relate to the development of addiction?

To address these questions, we present recent studies that confirm signaling of RPEs by DA neurons and discuss the possible neural bases for RPE computation. We then ask how the DA-RPE hypothesis might apply to addiction from both the theoretical and neurobiological perspective. Our position is that focused study of midbrain DAergic neurons and the neural networks in which they are embedded is required for complete understanding of DA both in normal learning and in addiction. We use the RPE model as a tool to formulate specific predictions regarding DA neural circuit function that can be tested using contemporary neurobiological techniques. With this goal in mind, we end with some suggestions of experiments that may clarify the role of DA-RPE's in addiction.

Reward-related activity in DA neurons

In a series of seminal studies, Schultz and colleagues reported that putative DA neurons in the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) respond to natural rewards, such as palatable food, with a burst of action potentials (Ljungberg et al., 1992; Schultz et al., 1993). Notably, the sign and magnitude of the DA neuron response is modulated by the degree to which the reward is expected (summarized in Bromberg-Martin et al., 2010; Schultz, 1998): surprising or unexpected rewards elicit strong increases in firing, whereas anticipated rewards produce little or no change. Conversely, if an anticipated reward fails to materialize, DA neuron firing is reduced below baseline. Thus DA neurons do not provide an invariant readout of the presence of a given reward. Rather, it appears that DA neurons encode an RPE, reporting the discrepancy between expected reward and reward that is obtained (Montague et al., 1996; Schultz et al., 1997). Another hallmark of the RPE-encoding DA neuron is the gradual transfer of neural activation from reward delivery to cue

onset during associative learning. Early in learning, when cue-reward associations are weak, DA neurons respond robustly to reward occurrence and weakly to reward-predictive cues. As learning progresses neural responses to the cue become pronounced and reward responses diminish. These canonical DA neuron responses to reward, and then later to the reward-predictive cue, have been observed in neural recordings from the VTA and SNc in primates (Bayer and Glimcher, 2005; Matsumoto and Hikosaka, 2009; Schultz et al., 1993), and rats (Pan and Hyland, 2005; Pan et al., 2005), and in identified DA neurons in the VTA of mice (Cohen et al., 2012).

RPEs matter because they are a central component of reinforcement learning theory which posits that increases in phasic DA to unpredicted reward are necessary for some types of reward learning, allowing an organism to predict the value of an upcoming event, thereby impacting subsequent decision-making and behavior (Glimcher, 2011; Schultz et al., 1997; Schultz and Dickinson, 2000). The relation of these ideas to reward-seeking behavior, including to drug-seeking behavior and addiction, is discussed in detail in the remainder of this review.

Other non-reward DA neuron signals and DA function beyond RPE signals

Salient non-reward stimuli are reported to induce DA neuron firing (Bromberg-Martin et al., 2010). Salience responses might reflect the potential behavioral importance of a stimulus, an inherently rewarding property of some salient stimuli, or generalization of neutral salient cues and reward cues when both are presented within contexts associated with high reinforcement probability (Kobayashi and Schultz, 2014). In addition, some DA neurons are activated by noxious stimuli (Brischoux et al., 2009; Matsumoto and Hikosaka, 2007). These neural responses to salient and aversive stimuli indicate that the concept of RPE may not encompass the entirety of phasic DA signals. Notably, different behavioral effects of phasic DA may arise due to activation of particular populations of DA neurons with distinct projection targets. For example, in both primates (Matsumoto and Hikosaka, 2009) and rodents (Lammel et al., 2011; Lammel et al., 2012), putative salience/aversion-signaling DA neurons are found more frequently in certain subregions of the SNc/VTA. In addition, DA neuron burst firing at different behaviorally-relevant times – for example, immediately before an action vs. upon delivery of unexpected reward – may have distinct effects, such as mediating aspects of motivation vs. learning, respectively. Finally, an individual phasic DA signal occurring at a specific time may be multiplexed (a cue-triggered burst might modulate ongoing behavior and contribute to future behavior via altered neuronal plasticity). Multiple means by which DA affects behavior have been proposed (for review, Berridge and Robinson, 2003; McClure et al., 2003; Niv et al., 2007; Robbins and Everitt, 2007; Salamone and Correa, 2012). Here we limit our discussion to the potential role of a DA neuron-mediated RPE in learning and addiction, and do not discuss other possible contributions of DA neuron firing to behavior.

Phasic DA signals as a biological implementation of RPE

Prediction error in formal models of learning

How do prediction errors aid learning? Theories of associative learning have long recognized that simply pairing a cue with reward is not sufficient for learning to occur. In addition to contiguity between two events, learning requires the subject to detect a discrepancy (RPE) between an expected reward and the reward that is actually obtained. This RPE acts as a teaching signal used to correct inaccurate predictions. In a highly influential model, Rescorla and Wagner (1972) proposed that a prediction error is computed at the end of individual learning trials (when the reward is delivered or withheld), causing predictions to be updated on a trial by trial basis. However, this separation of time into discrete trials is arbitrary and predictions can certainly be revised before the end of a trial as new information becomes available. In response to this limitation, the more recent temporal difference model of reinforcement (TD model), originating from computer science, assumes that predictions are generated and updated continuously (Sutton and Barto, 1998). At every moment, or state (S_t), a prediction error is computed by comparing the information available at that moment (be it an actual reward or a signal that reward is imminent) with what was predicted a moment before (in the previous state, S_{t-1}). The prediction error computed at a time t is thus defined by:

$$\text{Prediction error (t)} = R_t + V(S_t) - V(S_{t-1})$$

where R_t represents the value of the outcome present at a time t , and $V(S_t)$ and $V(S_{t-1})$ correspond to the value of the state t and $t-1$, determined by the predictions of upcoming reward computed in the state t and $t-1$, respectively. Because subjects value reward less as they become more distant in time, $V(S_t)$ is often corrected by a temporal discounting factor, omitted here for simplicity. This prediction error computed at time t is then used to update the predictive value of the preceding state:

$$V(S_{t-1})_{\text{new}} = V(S_{t-1})_{\text{old}} + \eta \text{Prediction Error}(t)$$

where η represents a learning rate parameter that determines the weight of prediction errors in the process of updating predictions. In addition, when the transition from one state to the next depends on a specific action, the prediction error resulting from this state transition can be used to update the value of the instrumental action:

$$\pi(A | S_{t-1})_{\text{new}} = \pi(A | S_{t-1})_{\text{old}} + \eta \text{Prediction error}$$

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

In addition to offering a “real time” account of predictions and prediction errors, an advantage of this model is that it accounts for the reinforcement of cues and actions that are temporally distant from the primary reward (credit assignment problem). Indeed, this model predicts that as the value of a cue (or state) that closely precedes the delivery of a reward increases, that cue can then itself produce an RPE and thereby reinforce more distal cues and actions. This back-propagation of RPE allows for identification of the earliest predictor of a primary reward and for the reinforcement of distal actions leading to an increase in the prospect of reward. The TD reinforcement learning model offers an elegant and

parsimonious computational account for many (but not all) learning phenomena related to changes in cue/state values (Miller et al., 1995).

How well do TD prediction errors parallel DA responses? We can see from Figure 1 that the TD reinforcement learning model can account for phasic DA responses in most situations. This striking parallel between theoretical concepts of prediction errors and DA responses led to the hypothesis that phasic DA signals represent the biological implementation of RPEs (Schultz et al., 1997). More specifically, it has been proposed that the error signal carried by phasic DA activation is broadcast to forebrain regions involved in reward learning, such as the striatum, where DA surges could function as teaching signals strengthening neural representations that facilitate reward receipt, possibly via alteration of the strength and direction of synaptic plasticity.

Evidence for DA teaching signals

The formulation discussed above describes behavioral learning with great precision but our interest here lies in whether the RPE model can help us understand how the brain can accomplish this learning. The parallel between phasic DA responses (recorded *in vivo*), and RPE (computed *in silicio*) suggests that, similarly to RPE, phasic DA responses constitute a teaching signal that reinforces stimuli and actions leading to positive outcome.

Studies using DA antagonists or selective DA lesions provide considerable evidence for DA's role in associative learning (for review, Costa, 2007; Wise, 2004). However, these techniques affect tonic as well as phasic DA levels at the time of training and often at the time of test which limits interpretation of the results in terms of DA-RPE acting as teaching signals (Costa, 2007). The advent of optogenetic approaches has allowed for more specific tests of DA-RPEs because temporally-precise and selective control of DA neurons can be achieved (Tsai et al., 2009; Witten et al., 2011). In transgenic animals expressing Cre recombinase under control of the tyrosine hydroxylase promoter (Th::Cre), cre-dependent viral vectors injected into the VTA/SNc can be used to induce expression of the light-sensitive sodium channel, channelrhodopsin-2 (ChR2), selectively in DA neurons. Using these optogenetic tools, it has been shown that rodents develop a preference for a compartment paired with phasic, but not tonic, optical activation of VTA DA neurons, and self-stimulate VTA or SNc DA neurons at high rates (Ilango et al., 2014; Steinberg et al., 2014; Tsai et al., 2009; Witten et al., 2011). These results are consistent with the hypothesis that phasic DA signals encode positive RPEs, acting as teaching signals.

The role of phasic DA signals in error-driven learning was most clearly evidenced by manipulating DA neuron activity in a blocking paradigm (Steinberg et al., 2013). The phenomenon of blocking is a powerful illustration of prediction errors in learning (Kamin, 1969). In this procedure, the acquisition of a cue-reward association is impaired (or blocked) if another cue, present in the environment at the same time, already signals reward delivery. For instance, consider two cues (A and X), presented simultaneously (i.e., in compound) and followed by reward delivery. Conditioning to one element of the compound (X) is reduced – blocked – if the other element (A) has already been established as a reliable predictor of the reward (Holland, 1999; Steinberg et al., 2013). This indicates that simple contiguity between a stimulus and reward is not sufficient for conditioning; learning requires the detection of a

discrepancy between expected and actual events. In such a blocking procedure, activity of DA neurons parallels RPEs (Waelti et al., 2001) providing compelling correlative support for phasic DA signals and error-based learning.

This notion was further tested using optical stimulation of VTA DA neurons within the blocking experimental design (Steinberg et al., 2013). In this study, stimulation was delivered at the time of reward receipt, but only during the compound stimulus phase when reward is fully predicted by the original training cue, and when learning about the new cue presented in compound normally does not occur. When paired with the delivery of the expected reward, phasic DA neuron stimulation produced an artificial RPE sufficient to drive learning about the normally blocked cue (Figure 2). Notably the effect was temporally-specific: equivalent stimulation in the intertrial interval not paired with reward failed to unblock learning. These results indicate that brief phasic stimulation of DA precisely timed with reward presentation can mimic a normally-absent RPE and rescue blocked learning. The role of phasic DA activity as an error correction signal was further demonstrated in a follow up experiment in which DA neuron photostimulation during the omission of expected reward interfered with negative RPE signaling (encoded by pauses in DA neurons) and attenuated behavioral extinction. Collectively these findings indicate that DA neurons encode RPE and drive error-correction learning, as in TD reinforcement models.

Computation of RPE in DA neurons

If DA neuron firing provides an RPE that drives learning, then it must be imbedded in a circuit that allows extraction of that error signal from neural elements encoding the expected and the actual reward. What might an RPE-generating circuit look like? Mapping the connectivity of midbrain DA neurons is critical for determining how RPEs may be generated. VTA and SNc DA neurons receive diverse excitatory and inhibitory inputs, ranging from the prefrontal cortex (PFC) to the brainstem (Sesack and Grace, 2010; Watabe-Uchida et al., 2012). In addition, VTA DA neurons (and to some extent, SNc) are intermingled with GABA and glutamate neurons, forming a complex local circuitry (Nair-Roberts et al., 2008; Tepper and Lee, 2007; Yamaguchi et al., 2015; Yamaguchi et al., 2013). These long- and short-range inputs, along with intrinsic cellular properties, shape DA neuronal firing.

The generation of phasic bursting in DA neurons - the RPE itself - depends on glutamatergic transmission (Grace et al., 2007). Genetic inactivation of glutamatergic NMDA receptors on DA neurons reduces phasic burst firing without affecting tonic firing (Wang et al., 2011; Zweifel et al., 2009). Similar results are observed after injection of NMDA receptor antagonists in the VTA/SNc (Chergui et al., 1993; Sombers et al., 2009). Multiple inputs contribute to DA bursting (Grace et al., 2007); here we focus on inputs for which we also have information regarding behavioral correlates.

The pedunculopontine tegmental nucleus (PPTN) is one important candidate via its glutamatergic (and cholinergic) inputs to DA neurons (Kobayashi and Okada, 2007). PPTN neurons show phasic responses to primary rewards and reward cues with shorter latency than DA neurons, and in projections to the lateral SNc, encode rewards and reward expectation (Hong and Hikosaka, 2014). This region also responds to sensory cues prior to

learning (Pan and Hyland, 2005) suggesting the PPTN is important for indicating the occurrence of stimuli. Recording of midbrain DA neurons following pharmacological inactivation of the PPTN confirmed that this region participates in cue-evoked bursts in DA neurons (Pan and Hyland, 2005).

Another important source of glutamatergic (and peptidergic) inputs to midbrain DA neurons, particularly in the VTA, is the lateral hypothalamus (LH) (Watabe-Uchida et al., 2012). The LH has long been associated with motivated behaviors. LH neurons respond to primary rewards and learned reward cues (Nakamura and Ono, 1986; Nieh et al., 2015; Rolls et al., 1976), electrical stimulation of LH neurons results in transient DA release in the NAc (Lee et al., 2014), and optogenetic stimulation of LH inputs to the VTA is reinforcing in rats and increases reward-seeking behavior (Kempadoo et al., 2013; Nieh et al., 2015). Of note, most reward-evoked activity in PPTN and LH neurons is not modulated by expectation of reward (Kobayashi and Okada, 2007; Nakamura and Ono, 1986; but see Hong and Hikosaka, 2014). Overall, evidence suggests that PPTN and LH contribute excitatory input critical to the computation of RPE in DA neurons by providing information for the reward prediction, $V(S_t)$ (i.e., the learned cues) and/or information about the reward itself (R_t).

In addition to excitation by glutamate, burst firing requires the reduction of inhibition onto DA neurons (Aggarwal et al., 2012; Grace et al., 2007; Kitai et al., 1999). DA neurons receive numerous GABAergic synaptic inputs; in the SN, 40–70% of synaptic inputs onto DA neurons are GABAergic (Henny et al., 2012; Tepper and Lee, 2007). Some of this GABAergic modulation may come from the LH. A recent study found that optogenetic stimulation of the LH-VTA GABAergic projection increased feeding (Nieh et al., 2015), in agreement with a second study that found photostimulation of LH GABAergic cell bodies also increases feeding and is reinforcing (Jennings et al., 2015). This GABAergic LH-VTA pathway may provide information about rewards that increases DA neuronal activity indirectly through a VTA GABAergic interneuron, although there is also evidence that this pathway directly innervates DA neurons (Nieh et al., 2015).

Burst firing may also be controlled by the striatum. It has been suggested that striatal GABAergic medium spiny neurons (MSNs) participate in the computation of RPE by orchestrating inhibition and disinhibition of DA neurons, providing inputs that may encode R_t and $V(S_t)$ (Aggarwal et al., 2012; Morita et al., 2012, 2013). MSNs encode both rewards and cue-induced reward predictions (Nicola et al., 2004; Roesch et al., 2009; Stalnaker et al., 2010) and project directly and indirectly onto DA neurons. It has been hypothesized that cue- or reward-evoked activation of D1 receptor-expressing (D1R) MSNs that project predominantly to GABAergic neurons in the VTA or SNr, facilitate burst firing of DA neurons in the VTA or SNc respectively, via a disinhibition mechanism (Morita et al., 2012). A recent study used optogenetic circuit manipulation *in vivo* to test this idea and confirmed that photostimulation of D1R MSNs activates VTA DA neurons via inhibition of VTA GABA neurons (Bocklisch et al., 2013). Hence this disinhibition pathway could be important for RPEs.

Central to the idea that DA neurons encode an RPE is the observation that reward-evoked activation progressively declines as reward becomes expected. However, unexpected

delivery of the same reward results in strong responding of DA neurons (Pan et al., 2005). Therefore, the neural inputs carrying excitatory information about the reward (R_t) have not decreased in efficacy over learning; instead this suggests that reward prediction results in synaptic inhibition on DA neurons that opposes the excitation produced by the reward itself. Reward predictions, for example triggered by the cue, thus may have dual and opposite influences on DA neurons: a rapid excitation followed by a delayed inhibition that peaks at the time of the expected reward and therefore can cancel reward-evoked excitation. Various models have suggested that this inhibition component is driven by MSNs in the patch compartment of the striatum directly projecting to DA neurons (e.g., Brown et al., 1999; Humphries and Prescott, 2010). However, the projection-specific activation afforded by optogenetics has revealed that stimulation of direct striatal inputs on DA neurons in the VTA or SNc evokes weak to no inhibitory currents in DA neurons (Bocklisch et al., 2013; Chuhma et al., 2011; Xia et al., 2011), indicating that these direct afferents are relatively silent. More recently, it has been proposed that the inhibitory component of reward prediction is driven by D2 receptor-expressing (D2R) striatal MSNs, reaching DA neurons in the SNc via the GPe, STN and SNr indirect pathway (Morita et al., 2012). A similar mechanism of indirect inhibition from D2R MSNs to VTA DA neurons may exist, involving VTA GABA neurons as the final relay (Groenewegen and Berendse, 1990; Morita et al., 2012, 2013). In agreement, using cell-type specific expression of ChR2, VTA GABA neurons were shown to encode reward prediction with a ramping of activity and to inhibit DA neurons, a potential means for participating in RPE computation (Cohen et al., 2012; Tan et al., 2012). Interestingly, the multiple relays that compose this indirect pathway could contribute to the time delay required for the computation of the temporal difference (Aggarwal et al., 2012; Morita et al., 2012, 2013); in this scenario, direct- as well as indirect-pathway MSNs could encode reward predictions, which is consistent with the observed synchronous activation of D1R and D2R MSNs during reward seeking (Cui et al 2013), but these predictions would contribute differentially to RPE computation by disinhibiting and inhibiting DA neurons, respectively. Consistent with this latter idea, optogenetic stimulation in mice of D1R or D2R MSNs contingent upon an action, increases or decreases, respectively, the probability of the animal engaging in that particular action (Kravitz et al., 2012). Of note, segregated cortical ensembles are proposed to innervate direct and indirect pathway MSNs thereby signaling the different states (S) from which predictions (V(S)) can be derived (Morita et al., 2012; Takahashi et al., 2011; Wilson et al., 2014).

What about the pause in firing in response to omission of an expected reward? A parsimonious explanation is that this pause in firing is driven by the same inhibitory mechanism that cancels reward-evoked excitation. In addition, reward omissions, as well as aversive outcomes, activate neurons in the lateral habenula (LHb) (Matsumoto and Hikosaka, 2007), which indirectly inhibits DA neurons via activation of GABAergic neurons in the rostromedial tegmentum nucleus (RMTg) (Hong et al., 2011; Zhou et al., 2009). A role for LHb in negative prediction errors was recently demonstrated by Tian and Uchida (in press) who examined neural responses of identified DA neurons in awake behaving mice after LHb lesions; they found a reduction both in the magnitude of the firing

rate decrease elicited by reward omission, and in the total number of units with a response to reward omission.

Using the information just described, a simplified circuit model for the calculation of RPEs by DA neurons can be constructed (Figure 3), in which LH and PPTN are proposed to critically contribute to R_t , and direct and indirect pathway MSNs drive the circuits that provide current and prior state values, respectively (Aggarwal et al., 2012; Morita et al., 2012). We now in principle have the neurobiological tools to test each element of this proposed circuit. For example, behavior-related firing of neuronal subtypes, such as DA neurons in the midbrain, or indirect or direct pathway neurons in the striatum, can be measured during learning using photo-tagging (Cohen et al., 2012) or deep brain calcium imaging (Jennings et al., 2015). The expression of opsins or DREADDS in neurons afferent to the DA or GABA neurons in the midbrain can permit excitation or inhibition in a projection-specific and temporally-restricted manner (as in Lammel et al., 2012; Mahler et al., 2014; Nieh et al., 2015; Tan et al., 2012; van Zessen et al., 2012) to test specific predictions of the model on learning and neural function. Identified synapses within this circuit can be probed *ex vivo* for learning-related changes. Note this circuit is undoubtedly a simplification; multiple inputs may be required for anyone one computational component in the TD equation, and individual inputs could partially compute an RPE; for example PPTN neurons may encode primary reward but a subset may encode reward prediction ($V(S_t) + R_t$; Hong and Hikosaka, 2014) or LHB may partially encode prediction error ($R_t - V(S_{t-1})$; Matsumoto and Hikosaka, 2007). In addition, this circuit could interact with other neural circuits computing similar error signals (Matsumoto and Hikosaka, 2007; McNally et al., 2011). Recent reports of critical distinctions among spatially or neurochemically different projections from the same brain region to the VTA/SNc will need to be considered (e.g., Hong and Hikosaka, 2014; Nieh et al., 2015). Current technological advances in measuring and manipulating genetically-defined cell types, including their neural projections, will allow testing and critical model updating.

DAergic projections and broadcasting of RPE

The neuronal and behavioral effects of DA neuron bursting necessarily depend upon which DA neurons fire and where they project. Midbrain DA neurons send efferent projections to many brain regions including the striatum, globus pallidus/ventral pallidum, subthalamic nucleus, thalamus, amygdala, BNST, prefrontal cortex, amygdala, and hippocampus (Beckstead et al., 1979; Gerfen et al., 1987; Swanson, 1982). The dopaminergic projections to the striatum are topographically arranged, such that DA neurons in the VTA project to ventro-medial striatum, while DA neurons in the ventral SNc project to dorso-lateral striatum (Haber et al., 2000; Ikemoto, 2007). DA projections are an integral part of the “spiraling” striatal circuits that emphasize a progressive ventral to dorsal recruitment that occurs via MSN feedback to DA neurons that innervate the neighboring striatal circuit (Haber et al., 2000; Ikemoto, 2007).

Based on widespread DA projections, it has been argued that DA release constitutes a global reinforcement signal, distributed throughout target regions to strengthen neural representations that facilitate reward receipt (Schultz, 1998). In agreement, DA neuron

activity and subsequent DA release have short- and long-term effects on neural circuits; these consequences on neuronal activity and plasticity are reviewed elsewhere (Gerfen and Surmeier, 2011; Kreitzer and Malenka, 2008; Lovinger, 2010; Pignatelli and Bonci, 2015; Wickens, 2009). Continued use of *ex vivo* measurements of synaptic changes in the context of natural and drug reward learning will help clarify how the phasic DA release of an RPE may induce plasticity that underlies learned associations governing reward seeking behavior.

Role of DA teaching signals in addiction

Learning to take drugs recruits associative learning processes. Subjects learn to associate environmental cues with drug availability and drug effects, and learn which specific action sequences are followed by drugs access. This is typically studied in preclinical models in which laboratory animals are trained to perform an instrumental response, such as a lever press, that is followed by a sensory cue, such as a brief auditory stimulus, that is predictive of intravenous (cocaine, amphetamine, morphine) or oral (alcohol) drug delivery. Animals learn to repeat the instrumental action to obtain more deliveries of drug. In other studies, Pavlovian learning is modeled by repeated pairing of a cue with drug delivery, without requiring the animal to emit an action to receive the drug. If a DA-RPE signal is sufficient to promote learning about natural reward cues (Steinberg et al., 2013, described above), is the DA-RPE signal sufficient for learning about drug cues? Likewise, if DA-RPE signals support action learning for natural reward outcomes, is the same true for drug outcomes?

A key point linking drugs of abuse to DA is that addictive drugs have direct and indirect pharmacological effects on DA neuronal activity and release thereby interacting with DA systems in a way that natural rewards do not. *In vivo* microdialysis studies reveal that cocaine, amphetamines, opiates, nicotine, cannabinoids and alcohol all, to varying degrees, increase extracellular DA in DA terminal regions, especially within the NAc and in some cases within the VTA itself (Di Chiara and Imperato, 1988; Nestler, 2005; Sulzer, 2011). Many studies have shown that self-administration of drugs of abuse depends largely (although not exclusively) on DA transmission (Koob and Volkow, 2010; Nestler, 2005; Pierce and Kumaresan, 2006; Wise, 2004). While these studies point to a general role of DA, the role of *phasic* DA signals in the development of addiction remains largely unknown. Indeed, prior studies necessarily used pharmacological and neurochemical approaches that produce long-lasting effects on DA transmission, which doesn't allow one to separate effects of basal DA levels resulting from tonic DA neuronal firing from effects of phasic firing. This distinction is important because it is the phasic DA responses that are implicated in learning (Steinberg et al., 2013; Tsai et al., 2009).

Thus, if drugs of abuse act like natural rewards upon DA systems to drive learning via a RPE mechanism, then their sensory or pharmacological attributes should activate bursting in DA neurons during learning like natural rewards do. To address this possibility, we can examine data obtained from *in vivo* fast scan cyclic voltammetry (FSCV), used to measure real-time rapid changes in striatal extracellular DA fluctuations ('transients') in relation to drug administration or various behavioral events. DA transients are dependent upon burst firing-induced release from DA neurons (Somers et al., 2009). In the awake animal, cocaine, amphetamine, ethanol, opioids and nicotine acutely increase DA transients in the

NAc (albeit with different potencies and time courses) and these increases are abolished by pharmacological suppression of DA neuron burst firing in the VTA (Cheer et al., 2007; Covey et al., 2014; Daberkow et al., 2013; Vander Weele et al., 2014; Wanat et al., 2009). Of note, cocaine was indeed found to increase DA neuron burst firing in awake rats (Koulchitsky et al., 2012).

Similarities and differences in food- vs. drug-related DA signals

When the reward is a sweet liquid or food pellet, reward-predictive cues and unexpected reward delivery elicit NAc DA increases time-locked to the behavioral events and expected reward does not elicit DA changes (Clark et al., 2013; Day et al., 2007; Flagel et al., 2011; Hart et al., 2014). Likewise, in animals trained to press a lever for food, DA release in the NAc is first elicited by response-contingent, but unexpected, food reward. As learning progresses and animals learn to predict food delivery, the phasic DA signal is transferred from the reward to cues that signal reward availability and trigger the instrumental response (Roitman et al., 2004), reflecting the backpropagation of RPE described in the TD reinforcement model. Self-initiated instrumental actions are also preceded by an increase in DA release (Wassum et al., 2012), perhaps caused by external cues or internal states unappreciated by the experimenter but causing a sudden increase in reward prediction and therefore an RPE (Wassum et al., 2012).

How does this compare to phasic fluctuations in DA measured in the NAc during drug self-administration? During cocaine self-administration, NAc DA also briefly increases prior to the instrumental response and in response to cues signaling drug delivery (Phillips et al., 2003; Stuber et al., 2005b). Though coincident with the lever press and drug infusion, the cue is the critical elicitor of this time-locked DA increase as presentation of the conditioned cue on its own is sufficient to induce a phasic DA increase (Phillips et al., 2003; Willuhn et al., 2012).

The critical difference between food and drug self-administration resides in the response to the reward itself. While food reward elicits a time-locked DA transient that progressively decreases as reward becomes expected (Figure 4A), the i.v. injection of cocaine elicits a delayed long-lasting increase in the frequency and amplitude of DA transients that is not clearly synchronized with specific behavioral events (Heien et al., 2005; Stuber et al., 2005a). Importantly, these drug-induced DA transients do not appear to be modulated by drug expectation (Figure 4B); they persist after the task is well-learned and the drug reward well-predicted (Stuber et al., 2005a; Willuhn et al., 2014). Although the overall basal frequency of DA transients appears to be altered by chronic cocaine, the acute effect of cocaine on transient changes in DA release is the same whether the animal or the experimenter administers the drug (Willuhn et al., 2014). As mentioned, other drugs of abuse also increase the amplitude and/or frequency of DA transients, suggesting they, too, would have this action after active self-administration.

Maladaptive prediction errors and addiction

In the framework of the TD reinforcement model, the expectation-independent surge of DA induced by drugs of abuse could contribute to drug-taking and, perhaps, addiction.

Specifically, it has been argued that drugs of abuse mimic a DA-RPE every time the drug is consumed (Di Chiara, 1999; Redish, 2004) via drug-induced increases in DA neuron firing and release. Over repeated drug use, the repetition of these DA signals would continue to reinforce drug-related cues and actions to pathological levels, biasing future decision-making towards drug choice. This is in contrast to natural rewards that produce error-correcting DA-RPE signals only until the predictions match the actual events. This viewpoint was expanded upon in a landmark paper which examined the outcomes of incorporating a drug-induced positive RPE signal in a TD learning computational model of addiction (Redish, 2004). In this model, every time the subject consumes a drug of abuse, a teaching signal is generated increasing the value of the state leading to drug receipt. Importantly, this drug-induced teaching signal is not dependent on the subject's predictions but set at a single value, representing the pharmacological effect of the drug on DA transmission. Therefore, the model predicts that with repeated drug administration, the value of states leading to drug receipt continues to increase without bounds eventually surpassing the value of states leading to natural reward (Figure 4C). As a result, when given the choice between drug and natural reward (more exactly, a choice between a state leading to drug and a state leading to natural reward), the subject develops a bias towards drug that strengthens with each drug use. In addition, this ever-increasing state value predicts decreasing elasticity over time such that experienced subjects are willing to emit considerable work for a given drug reward.

This model of 'persistent DA-RPE' provides a simple, yet powerful, explanation for several aspects of addiction. For example, it can explain the progressive allocation of more time and effort to drugs of abuse, often at the expense of natural reward and despite negative consequences (Deroche-Gamonet et al., 2004; Perry et al., 2013; Vanderschuren and Everitt, 2004; but see Lenoir et al., 2007).

A limitation of this model is that each drug administration is considered identical to the next in terms of consequences on DA transmission and error correction, independent of previous drug use history. This ignores the fact that drugs of abuse induce neuroadaptations in multiple brain regions and circuits, including the putative DA error correction circuit. Hence, as noted by Redish himself (2004), any model of addiction is incomplete without integration of the growing body of literature on the neurobiological effects of chronic drug-taking. In the following sections, we will discuss how the implementation of drug-induced DA-RPE in a biologically realistic circuit that integrates drug-induced neuroadaptations and anatomical considerations could extend our understanding of the role of DA-RPE in drug addiction and resolve some of the limitations of the original model.

Drug-induced neuroadaptations: consequences for RPE signaling

Drugs of abuse induce long-term neuroadaptations throughout the mesolimbic system, which durably alter the behavioral and neural effects of these drugs. A complete discussion of the nature of these drug-induced neuroadaptations can be found elsewhere (Hyman et al., 2006; Kalivas, 2009; Koob and Volkow, 2010; Luscher, 2013; Mameli and Luscher, 2011; Pignatelli and Bonci, 2015; Walker et al., 2015). We will focus here on changes in

excitatory and inhibitory transmission in the VTA and examine how these changes may affect striatal targets of DA neurons, as well as the consequences for associative learning.

Sensitization of DA responses

A well-studied adaptation to repeated administration of drugs of abuse is sensitization. Sensitization is observed as the progressive increase in a behavior, often locomotion, following repeated administration of the same dose of drug (Robinson and Becker, 1986; Vanderschuren and Kalivas, 2000). Behavioral sensitization induced by cocaine is accompanied by long-lasting increases in AMPA receptor (AMPA-R) mediated excitatory post-synaptic currents (EPSCs) onto VTA DA neurons (Ungless et al., 2001). In fact, exposure to drugs including morphine, nicotine, and alcohol, all increase AMPA-R-mediated currents on VTA DA neurons, an effect not observed with non-addictive drugs (Saal et al., 2003). This form of plasticity likely results in enhanced activation of DA neurons by glutamatergic input. Active cocaine self-administration similarly potentiates glutamatergic synapses on VTA DA neurons for at least 3 months (Chen et al., 2008). This effect is also observed after self-administration of food reward, although the increase only lasts 7 days (Chen et al., 2008). Therefore, transitory changes after natural reward may reflect normal learning, whereas longer-lasting changes after cocaine self-administration indicate an amplification of normal learning processes. Notably, increases in excitatory input onto DA neurons are observed after repeated experimenter-injected cocaine in the mouse in NAc-projecting DA neurons, but not in PFC-projecting DA neurons in the VTA, and not in DA neurons of the SNc (Lammel et al., 2011). Repeated exposure to cocaine also reduces the strength of inhibitory inputs on DA neurons, perhaps via an increase in long-term depression (LTD) of these GABAergic synapses (Liu et al., 2005; Pan et al., 2008) and/or the long-term potentiation (LTP) of GABA transmission between striatal D1R MSNs and VTA GABA neurons (Bocklisch et al., 2013).

The cocaine-induced combination of potentiated excitatory and reduced inhibitory inputs on DA neurons results in increased excitability of DA neurons (Bocklisch et al., 2013; White et al., 1995), which may be the basis for larger DA transients in the NAc in sensitized animals in response to electrical stimulation (Muscat et al., 1996; Ng et al., 1991; Williams et al., 1995). In addition, although the effect of drug sensitization on cue- or reward-evoked DA transients has not been reported, studies using *in vivo* microdialysis found that drug sensitization increases NAc DA responses to drugs, drug-cues, and natural reward (Addy et al., 2010; Fiorino and Phillips, 1999; Kalivas and Duffy, 1990; Williams et al., 1995). This potentiated DA response is associated with facilitated acquisition of an instrumental response for addictive drugs or natural rewards (Ferrario and Robinson, 2007; Horger et al., 1992; Horger et al., 1990; Olausson et al., 2006; Piazza et al., 1989; Taylor and Jentsch, 2001; Vezina et al., 1999). While this effect is often interpreted as sensitization of motivational or incentive properties of rewarding stimuli (Robinson and Berridge, 2001), it could also be interpreted as sensitization of reinforcing properties (i.e., ability to act as a teaching signal) of rewarding drugs and their associated cues (Redish, 2004; Zernig et al., 2007). According to this last hypothesis, drugs of abuse not only mimic RPE via their acute effect on DA transmission, but also potentiate future drug- and cue-induced RPE via long-lasting sensitization of mesolimbic DA circuits. As a consequence, the value of drug-related

cues and actions would not follow a linear progression, but an accelerated increase (Figure 4C). It is not known which glutamatergic inputs upon DA neurons undergo potentiation; in the context of the current discussion, identifying the enhanced input(s) would be of interest when considering possible drug-induced increases in the information ($V(S_t)$, R_t) they might carry.

Drug Neuroadaptations in the NAc – a consequence of DA-RPE?

Cocaine self-administration also induces long-lasting input-specific increases in excitatory synaptic transmission onto NAc MSNs (Lee et al., 2013; Ma et al., 2014; Pascoli et al., 2014). Using ChR2-assisted mapping to activate specific afferents to identified D1R or D2R MSNs, changes at identified synapses can be probed *ex vivo* after cocaine self-administration. These changes are candidate substrates for learned processes that underlie addictive-drug seeking behavior. In a recent example, enhanced AMPA-R EPSCs were detected at excitatory inputs from the PFC and the hippocampus to NAc shell D1R MSNs one month after cocaine self-administration. Reversing these experience-dependent synaptic changes via an optogenetic photostimulation LTD protocol altered cocaine seeking-behavior in an input-specific manner (Pascoli et al., 2014) strengthening the notion that these drug-induced synaptic changes reflect critical experience-dependent neuronal changes that underlie drug-seeking behavior. If a DA-RPE were required in some way for the formation of these experience-dependent synaptic alterations, then this finding would support a role for the DA-RPE as a teaching signal, driving later behavioral responses to drug-associated cues and contributing to propensity for drug-seeking actions. At the present time, it is not clear if these NAc changes induced by cocaine self-administration require DA action at the synapse for their formation; unlike the dorsal striatum, LTP and LTD of excitatory inputs onto MSNs in NAc slices is not DA-dependent (Fields et al., 2007; Lovinger, 2010; Nicola et al., 2000). Interestingly, NAc cocaine-induced excitatory synaptic plasticity does appear to be serially-dependent on initial cocaine-induced synaptic plasticity in VTA DA neurons (Mameli et al., 2009), and this VTA plasticity is DA dependent (Schilstrom et al., 2006). Note that DA neuron co-transmitters (peptides, glutamate) may be released during DA neuronal burst firing and these co-released transmitters may be requisite for DA-RPE effects on learning and behavior. Overall, there is a large amount of evidence for long-lasting drug-induced alterations in synaptic function, both at the level of the VTA and the NAc.

An intriguing shorter form of plasticity in the NAc core, consisting of increases in excitatory inputs to MSNs and increases in MSN spine diameter, is triggered by cocaine-associated cues and cocaine itself when these events lead to relapse in an animal model. This short-term plasticity is DA-dependent and is correlated with the behavior, reversing within hours of the relapse test (Gipson et al., 2013; Shen et al., 2014). This shorter form plasticity could be an example of the effect of DA-RPE signals on action selection, a possibility that remains to be tested. Of note, because some D1R MSNs project back to the VTA, synapsing upon GABA neurons (Xia et al., 2011), these short- and long-term striatal synaptic changes may feedback into altered computation of subsequent RPEs.

Regional progression of DA-RPEs in addiction

As mentioned, electrophysiological recordings found RPE-like signals in VTA as well as SNc DA neurons suggesting that this signal is a broadly-propagating global reinforcement signal distributed throughout the brain. However, more recent studies using FSCV to directly measure phasic DA signals revealed that DA-RPEs are not equally present throughout the striatum. In naïve untrained rats, unpredicted food delivery evokes phasic DA increases in the NAc core but surprisingly not in the dorsomedial striatum (DMS) or dorsolateral striatum (DLS) (Brown et al., 2011).

One possible explanation for this discrepancy between the electrophysiological and FSCV data is differential dorsal and ventral striatal modulation of DA levels at the synapse by variation in DA transporter density, resulting in regional variation in reuptake (Garris and Wightman, 1994), and by modulation of release by cholinergic inputs on DA terminals (Threlfell and Cragg, 2011), both of which could alter synaptic DA. A second possibility is that the discrepancy is due to differences in the amount of training subjects typically receive in these experiments. In electrophysiological recording studies with monkeys, subjects are generally well-trained, while rats used in FSCV experiments often have limited task experience. Therefore, it is possible that extended training results in activation of additional DA neurons including initially non-responsive neurons in the SNc. This hypothesis suggests that region-selective distribution of RPE DA signals is not fixed but varies with training history. Consistent with this idea, delivery of unpredicted food reward does evoke a phasic DA signal in the DMS after subjects learn to associate cue presentation with reward, in contrast to its lack of effect in eliciting DMS DA before training (Brown et al., 2011).

The underlying mechanism for regional progression of DA transients, and perhaps DA-RPEs, over the course of training is not known but may involve direct descending projections of striatal D1R MSNs onto GABAergic midbrain neurons. Indeed, we suggested earlier that these projections are involved in the computation of RPE by disinhibiting DA neurons in response to reward cues (Figure 3). An important feature of these projections is their non-reciprocal organization in the form of open “spiraling” loops connecting the more ventromedial to the more dorsolateral striatal regions (Figure 5A). This anatomical circuitry must be taken into account when formulating hypotheses of DA-RPEs in learning. In this framework, reward-evoked DA-RPEs initially restricted to the NAc contribute to increase the value of the cue (or state) that consistently precedes reward delivery. As the value of the cue increases, its presentation activates direct pathway MSNs in the NAc which act to disinhibit ventral but also more lateral DA neurons in the midbrain, provoking cue-evoked DA-RPEs in the NAc, but now also, the DMS. Generally speaking, the architecture of this circuit implies that the temporal backpropagation of DA-RPE from the reward itself to its earliest predictor also involves a regional propagation of DA-RPE from the more ventromedial to more dorsal and lateral striatal domains. In this way, over time, multiple DA-RPEs could be computed simultaneously within multiple (although not entirely independent) striatal subcircuits coursing through the NAc and the dorsal striatum.

How do drugs of abuse affect the regional progression/propagation of DA-RPEs? As previously mentioned, cocaine potentiates D1R MSN GABAergic inputs on midbrain GABAergic neurons (Bocklisch et al., 2013). The topology of this pathway suggests that

this potentiation facilitates the disinhibition of DA neurons in the VTA as well as the SNc, thereby possibly accelerating and extending the normal propagation of RPEs from NAc to more dorsal and lateral striatal subregions (Figure 5B). Hence, a state value ($V(S_t)$) signal from one striatal circuit may be propagated to the next. Consistent with this idea, a recent study showed that in animals with limited cocaine self-administration, the presentation of a cue that signals drug delivery evokes DA release in the NAc but not DLS. However, with a prolonged history of cocaine self-administration, cue-evoked responses progressively appear in the DLS while simultaneously decreasing in the NAc (Willuhn et al., 2012). These findings are in agreement with prior measurements of DA using *in vivo* microdialysis, which found that DA increases to response-contingent cocaine cues were detected in the dorsal striatum but not NAc in well-trained rats self-administering cocaine (Ito et al., 2000, 2002).

While the contribution of drug-induced potentiation of D1R MSN striatal inputs on midbrain GABA neurons to the ventral-to-dorsal RPE progression is at this point a hypothesis to be tested, there are experimental findings consistent with a requisite role of the NAc in initiating the propagation of RPEs to other striatal domains. Remarkably, lesions of the NAc prior to cocaine self-administration prevent the development of cue-evoked FSCV DA signals in the DLS (Willuhn et al., 2012). Although the progressive recruitment of more lateral midbrain DA neurons for the encoding of RPEs may depend initially on direct descending projections from the NAc, a form of Hebbian plasticity can then occur in the VTA and SNc, allowing excitatory sensory inputs from the PPTN, LH, or superior colliculus, for example, to progressively gain control of more lateral DA neurons.

What are the behavioral consequences of this regional progression/propagation of DA-RPEs? Functionally, the progression of DA-RPEs to extended striatal domains allows DA-RPE signals to promote plasticity in different corticostriatal circuits and thereby engage different psychological processes in the control of behavior (Everitt and Robbins, 2005; Porrino et al., 2004; Yin and Knowlton, 2006). Neural activity and plasticity in different striatal regions supports different forms of associative learning. For instance, pharmacological inactivation and neural imaging studies show that the NAc is involved in Pavlovian predictions (O'Doherty et al., 2004; Robbins and Everitt, 1992). In contrast, the DLS is involved with instrumental stimulus-response (S-R) associations whose expression is independent of the representation of the reward and therefore resistant to post-conditioning changes in outcome value (Robbins and Everitt, 1992; Yin et al., 2004). Thus, by potentially accelerating the recruitment of DA neurons in the SNc for the encoding of RPEs and thereby accelerating the emergence of DA teaching signals in the DLS, cocaine exposure could promote the acquisition of S-R associations resulting in habitual behavior. Consistent with this proposed mechanism, repeated exposure to cocaine or amphetamine accelerates habit formation (Corbit et al., 2014a; LeBlanc et al., 2013; Nelson and Killcross, 2006; Schoenbaum and Setlow, 2005).

Evidence also supports a role for more dorsal striatal regions as drug-seeking itself becomes habitual for cocaine and alcohol (Corbit et al., 2012, 2014b; Zapata et al., 2010). Involvement of the full ventral-to-dorsal spiral in cocaine-seeking is supported by a circuit disconnection study in which an NAc lesion in one hemisphere and DLS microinfusion of a

DA antagonist in the other hemisphere was shown to disrupt this behavior (Belin and Everitt, 2008). Finally, in Rhesus monkeys self-administering cocaine, increased involvement of more dorsal striatal circuits was demonstrated, using metabolic measures, after one hundred sessions as compared to five sessions of cocaine self-administration, in agreement with a progressive recruitment along the ventral-to-dorsal axis (Porrino et al., 2004).

In humans, the finding that alcohol cues are associated with greater activation of dorsal striatal regions in heavy drinkers and greater activation of ventral striatal regions in light drinkers is also supportive of a ventral to dorsal change (Vollstadt-Klein et al., 2010). Preferential dorsal striatal activation (Garavan et al., 2000) and increased dorsal striatal DA release (Volkow et al., 2006; Wong et al., 2006) in response to cocaine cues in human cocaine addicts has also been observed.

Note that the progressive involvement of the striatonigral complex during instrumental conditioning and the development of S-R habits is a natural process observed after extended training in many tasks involving natural reward (Faure et al., 2005; Yin et al., 2004) and is not necessarily equivalent with the development of compulsion. It has been suggested that drug-induced neuroadaptations subvert the neural mechanisms of habit formation (possibly via the mechanism described above) causing exaggerated, maladaptive S-R associations (Belin et al., 2009; Everitt and Robbins, 2005). Such pathological learning could be responsible for compulsive drug-seeking by compelling addicts to seek drug in response to certain stimuli, regardless of potential adverse consequences. However, a recent study found that phasic DA signals correlated with cocaine-seeking and a cocaine-predictive cue are greatly reduced in both the NAc and the DLS in animals whose drug-taking accelerated over many days (Willuhn et al., 2014). This acceleration in intake is often taken as a model of excessive or compulsive cocaine use; these results are in-line with the notion that regional progression of DA-RPE is important for the earlier stages of drug-seeking behavior (and perhaps again following withdrawal, when drug-induced DA decrements have resolved), but may not be required for chronic maintenance of addictive-like behavior, a question to be resolved by future studies.

Computational models of DA-RPE, DAergic spirals and addiction

The ideas expressed above are captured by a recent computational model that incorporates the TD model of reinforcement by DA neurons into a biologically-realistic corticostriatal architecture (Keramati and Gutkin, 2013). Inspired by literature on the hierarchical organization of behavior (Botvinick et al., 2009; Lashley, 1951; Timberlake, 1994) and prior implementations of the actor-critic model in addiction (for example, Takahashi et al., 2008), the authors argue that a given choice can be broken down into several nested subroutines, from a more abstract motivational level (ex: the desire for a cigarette), to more detailed motor commands involved in the execution of this desire (reach for packet of cigarette, reach for lighter, etc.). Neurobiologically, the authors propose that different levels of abstraction are implemented by corresponding corticostriatal circuits, with DA-RPEs in the NAc reinforcing the more coarse-grained motivational levels of a decision while DA-RPEs in the DLS reinforce the value of finer-grained lower-level actions, with multiple

corticostratial circuits (and degrees of abstraction) between these extremes. The interactions between different levels of reinforcement are proposed to be mediated by MSN feedback on DA neurons.

Intriguingly, simulations conducted by Keramati and Gutkin (2013) with this model find different outcomes for natural and drug reward. While instrumental conditioning with natural reward results in consistent valuation across different degrees of abstraction, conditioning with drugs of abuse results in inconsistencies in valuation, with the low level discrete drug-seeking actions reaching a significantly higher value than the higher-level decision to engage into drug-seeking. Thus, the value of engaging in drug use can be relatively low at motivational or cognitive levels, but extremely high at the level of action selection (Keramati and Gutkin, 2013). Interestingly, the predicted DA transients in NAc and DLS from this model parallel those observed to a cocaine cue in Willuhn et al. (2012).

This framework for DA-RPEs and addiction may therefore address an additional shortcoming of the original model of persistent DA-RPEs that, in its simplest formulation, fails to capture the ambivalence addicts experience towards drugs. According to the model, drug users choose to engage in drug consumption because it is, among other alternatives present at that time, the action of highest value. In this scenario, each drug choice is always, from the perspective of the addict, the best and more rational choice to make. However this is not the experience of many treatment-seeking addicts (Heather, 1998; Skog, 2003). Despite a strong bias towards drugs, addicts can often acknowledge that drug use is not the best decision (Stacy and Wiers, 2010). Discordant valuation signals at different levels of the corticostratial system may contribute to these inconsistencies.

Remaining questions

The DA-RPE hypothesis provides a compelling framework that continues to inspire reward and addiction research. Here we have illustrated how integration of this hypothesis with neural circuit models and new experimental findings provides us with a way forward in understanding more deeply the role of DA in addiction. The evidence for a role of DA-RPE in normal reward learning, at least of some forms (Flagel et al., 2011), is compelling, with correlated measures of DA neuron spiking (Cohen et al., 2012; Glimcher, 2011; Waelti et al., 2001) and DA release (Clark et al., 2013; Day et al., 2007; Hart et al., 2014) supporting such a role. As well, learning can be driven by production of an artificial DA-RPE by ChR2-mediated activation of VTA DA neurons at the time of expected reward (Steinberg et al., 2013). These findings must be brought to the circuit and synaptic level for a mechanistic understanding. An outline of a possible RPE-generating circuit has been described that can be tested through selective manipulation of its components during learning. In addition, one can test whether DA release specifically during the reward or cue predictor modulates or modifies specific synaptic changes that underlie later emitted behavior.

Because addictive behavior likewise involves the acquisition of associations between cues and rewards and actions and reward, DA-RPEs logically should also contribute to the acquisition of drug-seeking and taking behavior. However, because drugs of abuse stimulate DA neurons via their pharmacological properties, additional acute and long-term alterations

of the relevant neural circuitry occur. The distinct pharmacokinetics and mechanisms of action of different drugs of abuse undoubtedly have implications for the proposed RPE-like signal and learning processes that are proposed here and remain an important area of future investigation. The outstanding question is how these acute and long-term pharmacological effects of drugs alter the ongoing and future probability of drug-cue responsiveness and of actions made to obtain drug. Does the ability of drugs of abuse to stimulate DA produce aberrant DA-RPE-mediated learning processes within and across striatal circuits that underlie addictive behavior? Is this mediated directly through the persistent ability of drugs to stimulate DA activity, acting as an aberrant RPE signal? Or is this mediated through drug-induced changes in the mesolimbic and nigrostriatal circuits that serve to pathologically amplify the response to other stimuli that induce DA-RPEs, such as cue drug predictors (or both)?

These questions are not as yet answered. What we know at the present time includes the following: 1) drugs of abuse increase phasic DA release, as do natural rewards, fulfilling one requirement for acting as a DA-RPE; 2) during cocaine self-administration, phasic DA release is observed in DA terminal regions in response to cocaine-paired cues, as well as cocaine; 3) drugs of abuse produce both short- and long-lasting changes in synaptic plasticity within the putative error computation circuit, altering excitatory and inhibitory transmission at specific inputs onto VTA DA and GABA neurons and onto striatal MSNs; 4) synapse-specific reversal of some of these long-lasting changes in synaptic plasticity reduces or otherwise alters behavior in animal models of drug seeking.

These findings are exceptionally promising, and point directly to additional studies to address the possible function of DA-RPE in aspects of addiction:

1. First, if drugs of abuse act like natural rewards on DA systems to drive learning via an RPE mechanism, then their sensory or pharmacological attributes should activate bursting in DA neurons *during learning* like natural rewards do. While current findings suggest this is so, drug-induced DA neuron activation during early learning must be demonstrated to exist. If these DA signals that accompany drug injection serve as an RPE that drives learning, temporally-selective blockade of this signal should block learning that is integral to the development of addiction. To address these points, measurements of the activity of identified DA neurons using electrophysiology, deep brain calcium imaging, and FSCV can be made to determine when these neurons are activated and when DA is released relative to behavioral and drug events during learning, and temporally-specific inhibition of DA neuron activity can be conducted to counter the drug-induced putative DA-RPE to test causality. The feasibility of this latter approach was recently demonstrated in Th::Cre rats in which the light-sensitive inhibitory channel, halorhodopsin (NpHR), was expressed in DA neurons and NpHR-mediated inhibition of VTA DA somata was found to greatly reduce cocaine-induced phasic DA release in the NAc (McCutcheon et al., 2014).
2. Next, do drug-evoked synaptic changes in the putative DA-RPE circuit contribute to enhancement of the RPE signal itself, including the accelerated transfer of the RPE signal from ventral to more dorsal striatal circuits? Blockade or reversal of

these changes in a precise synapse-specific manner, coupled with electrophysiological or FSCV measures of DA neuron activity at different levels of the striatal circuitry can address this question.

3. Finally, do the circuit manipulations that affect the putative DA RPE signal correspondingly alter behavior? For example, if the emergence of DA-RPEs in the DLS depends upon the drug-induced potentiation at excitatory synapses onto NAc MSNs (Ma et al., 2014; Pascoli et al, 2014) or of the D1R MSN projection to VTA GABA neurons (Bocklisch et al., 2013), then the inhibition of plasticity selectively at these synapses should prevent regional propagation of DA-RPE and thereby prevent development of addiction-related behavior and/or compulsive drug-seeking. Taking the opposite approach, using optogenetics to mimic a DA-RPE in the DLS during a task involving natural reward should accelerate the development of habitual and perhaps compulsive reward-seeking.

Alternative views

We have focused our review on one aspect of DA neuron function and its possible contribution to addiction. This is not to give short shrift to other critical neural changes that accompany drug abuse, such as the well-documented decrements in prefrontal cortical function found in human addicts that surely contribute to maladaptive decision-making, and the adaptations in brain stress systems that may also critically contribute to compulsive drives to use drugs (Koob and Le Moal, 2008; Koob and Volkow, 2010). Multiple alterations in neural circuits likely interact to contribute to the disease state of addiction.

Importantly, it is possible that the persistent DA signal will be identified as a key component of the addictive process, but differently than focused upon here. For example, a prominent view is that the DA surge produced by drugs of abuse sensitizes incentive motivational processes, including those driven by learning, rather than directly providing a teaching signal (Robinson and Berridge, 2001). It is also possible that drug-induced DA increases will be shown to directly reinforce behavioral responses to cues and actions, but not to contribute to mechanisms that drive these behavioral responses to compulsive levels seen in the addict. Thus, in the end, as new results continue to emerge, theories may be updated, interpretations may change.

Conclusion

In conclusion, the importance of a DA-RPE in addiction remains an intriguing open question. What this consideration has highlighted is the remarkable progress made in recent times by investigating DA function at the level of circuits and synapses, taking advantage of contemporary tools allowing for spatial and temporal precision. This circuit-based approach combined with sophisticated animal models of well-defined addiction phenotypes is required to address the precise role(s) of DA in addiction.

ACKNOWLEDGMENTS

Supported in part by US National Institutes of Health grants DA035943, AA018025, AA014925. We thank members of our lab for input on this manuscript.

References

- Addy NA, Daberkow DP, Ford JN, Garris PA, Wightman RM. Sensitization of rapid dopamine signaling in the nucleus accumbens core and shell after repeated cocaine in rats. *J. Neurophysiol.* 2010; 104:922–931. [PubMed: 20554845]
- Aggarwal M, Hyland BI, Wickens JR. Neural control of dopamine neurotransmission: Implications for reinforcement learning. *Eur. J. Neurosci.* 2012; 35:1115–1123. [PubMed: 22487041]
- Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron.* 2005; 47:129–141. [PubMed: 15996553]
- Beckstead RM, Domesick VB, Nauta WJ. Efferent connections of the substantia nigra and ventral tegmental area in the rat. *Brain Res.* 1979; 175:191–217. [PubMed: 314832]
- Belin D, Everitt BJ. Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron.* 2008; 57:432–441. [PubMed: 18255035]
- Belin D, Jonkman S, Dickinson A, Robbins TW, Everitt BJ. Parallel and interactive learning processes within the basal ganglia: Relevance for the understanding of addiction. *Behav. Brain Res.* 2009; 199:89–102. [PubMed: 18950658]
- Berke JD, Hyman SE. Addiction, dopamine, and the molecular mechanisms of memory. *Neuron.* 2000; 25:515–532. [PubMed: 10774721]
- Berridge KC. From prediction error to incentive salience: Mesolimbic computation of reward motivation. *Eur. J. Neurosci.* 2012; 35:1124–1143. [PubMed: 22487042]
- Berridge KC, Robinson TE. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Res. Rev.* 1998; 28:309–369. [PubMed: 9858756]
- Berridge KC, Robinson TE. Parsing reward. *Trends Neurosci.* 2003; 26:507–513. [PubMed: 12948663]
- Bocklisch C, Pascoli V, Wong JC, House DR, Yvon C, Roo M, Tan KR, Luscher C. Cocaine disinhibits dopamine neurons by potentiation of GABA transmission in the ventral tegmental area. *Science.* 2013; 341:1521–1525. [PubMed: 24072923]
- Botvinick MM, Niv Y, Barto AC. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition.* 2009; 113:262–280. [PubMed: 18926527]
- Brischoux F, Chakraborty S, Brierley DI, Ungless MA. Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:4894–4899. [PubMed: 19261850]
- Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron.* 2010; 68:815–834. [PubMed: 21144997]
- Brown HD, McCutcheon JE, Cone JJ, Ragozzino ME, Roitman MF. Primary food reward and reward-predictive stimuli evoke different patterns of phasic dopamine signaling throughout the striatum. *Eur. J. Neurosci.* 2011; 34:1997–2006. [PubMed: 22122410]
- Brown J, Bullock D, Grossberg S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J. Neurosci.* 1999; 19:10502–10511. [PubMed: 10575046]
- Cheer JF, Wassum KM, Sombers LA, Heien ML, Ariansen JL, Aragona BJ, Phillips PE, Wightman RM. Phasic dopamine release evoked by abused substances requires cannabinoid receptor activation. *J. Neurosci.* 2007; 27(4):791–795. [PubMed: 17251418]
- Chen BT, Bowers MS, Martin M, Hopf FW, Guillory AM, Carelli RM, Chou JK, Bonci A. Cocaine but not natural reward self-administration nor passive cocaine infusion produces persistent LTP in the VTA. *Neuron.* 2008; 59:288–297. [PubMed: 18667156]
- Chergui K, Charléty PJ, Akaoka H, Saunier CF, Brunet JL, Buda M, Svensson TH, Chouvet G. Tonic activation of NMDA receptors causes spontaneous burst discharge of rat midbrain dopamine neurons in vivo. *Eur. J. Neurosci.* 1993; 5:137–144. [PubMed: 8261095]
- Chuhma N, Tanaka KF, Hen R, Rayport S. Functional connectome of the striatal medium spiny neuron. *J. Neurosci.* 2011; 31:1183–1192. [PubMed: 21273403]
- Clark JJ, Collins AL, Sanford CA, Phillips PE. Dopamine encoding of pavlovian incentive stimuli diminishes with extended training. *J. Neurosci.* 2013; 33:3526–3532. [PubMed: 23426680]

- Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*. 2012; 482:85–88. [PubMed: 22258508]
- Corbit LH, Chieng BC, Balleine BW. Effects of repeated cocaine exposure on habit learning and reversal by n-acetylcysteine. *Neuropsychopharmacology*. 2014a; 39:1893–1901. [PubMed: 24531561]
- Corbit LH, Nie H, Janak PH. Habitual alcohol seeking: Time course and the contribution of subregions of the dorsal striatum. *Biol. Psychiatry*. 2012; 72:389–395. [PubMed: 22440617]
- Corbit LH, Nie H, Janak PH. Habitual responding for alcohol depends upon both AMPA and D2 receptor signaling in the dorsolateral striatum. *Front Behav Neurosci*. 2014b; 8:301. [PubMed: 25228865]
- Costa RM. Plastic corticostriatal circuits for action learning: What's dopamine got to do with it? *Ann. N. Y. Acad. Sci.* 2007; 1104:172–191. [PubMed: 17435119]
- Covey DP, Roitman MF, Garris PA. Illicit dopamine transients: Reconciling actions of abused drugs. *Trends Neurosci*. 2014; 37:200–210. [PubMed: 24656971]
- Daberkow DP, Brown HD, Bunner KD, Kraniotis SA, Doellman MA, Ragozzino ME, Garris PA, Roitman MF. Amphetamine paradoxically augments exocytotic dopamine release and phasic dopamine signals. *J. Neurosci*. 2013; 33:452–463. [PubMed: 23303926]
- Day JJ, Roitman MF, Wightman RM, Carelli RM. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci*. 2007; 10:1020–1028. [PubMed: 17603481]
- Deroche-Gamonet V, Belin D, Piazza PV. Evidence for addiction-like behavior in the rat. *Science*. 2004; 305:1014–1017. [PubMed: 15310906]
- Di Chiara G. Drug addiction as dopamine-dependent associative learning disorder. *Eur. J. Pharmacol*. 1999; 375:13–30. [PubMed: 10443561]
- Di Chiara G, Imperato A. Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proc. Natl. Acad. Sci. U. S. A.* 1988; 85:5274–5278. [PubMed: 2899326]
- Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nat. Neurosci*. 2005; 8:1481–1489. [PubMed: 16251991]
- Faure A, Haberland U, Conde F, El Massioui N. Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *J. Neurosci*. 2005; 25:2771–2780. [PubMed: 15772337]
- Ferrario CR, Robinson TE. Amphetamine pretreatment accelerates the subsequent escalation of cocaine self-administration behavior. *Eur. Neuropsychopharmacol*. 2007; 17:352–357. [PubMed: 17029912]
- Fields HL, Hjelmstad GO, Margolis EB, Nicola SM. Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annu. Rev. Neurosci*. 2007; 30:289–316. [PubMed: 17376009]
- Fiorino DF, Phillips AG. Facilitation of sexual behavior and enhanced dopamine efflux in the nucleus accumbens of male rats after d-amphetamine-induced behavioral sensitization. *J. Neurosci*. 1999; 19:456–463. [PubMed: 9870973]
- Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PE, Akil H. A selective role for dopamine in stimulus-reward learning. *Nature*. 2011; 469:53–57. [PubMed: 21150898]
- Garavan H, Pankiewicz J, Bloom A, Cho JK, Sperry L, Ross TJ, Salmeron BJ, Risinger R, Kelley D, Stein EA. Cue-induced cocaine craving: Neuroanatomical specificity for drug users and drug stimuli. *A. J. Psychiatry*. 2000; 157:1789–1798.
- Garris PA, Wightman RM. Different kinetics govern dopaminergic transmission in the amygdala, prefrontal cortex, and striatum: An in vivo voltammetric study. *J. Neurosci*. 1994; 14:442–450. [PubMed: 8283249]
- Gerfen CR, Herkenham M, Thibault J. The neostriatal mosaic: Ii. Patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. *J. Neurosci*. 1987; 7:3915–3934. [PubMed: 2891799]
- Gerfen CR, Surmeier DJ. Modulation of striatal projection systems by dopamine. *Annu. Rev. Neurosci*. 2011; 34:441–466. [PubMed: 21469956]

- Gipson CD, Kupchik YM, Shen H, Reissner KJ, Thomas CA, Kalivas PW. Relapse induced by cues predicting cocaine depends on rapid, transient synaptic potentiation. *Neuron*. 2013; 77:867–872. [PubMed: 23473317]
- Glimcher PW. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:15647–15654. [PubMed: 21389268]
- Grace AA, Floresco SB, Goto Y, Lodge DJ. Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends Neurosci.* 2007; 30(5):220–227. [PubMed: 17400299]
- Groenewegen HJ, Berendse HW. Connections of the subthalamic nucleus with ventral striatopallidal parts of the basal ganglia in the rat. *J. Comp. Neurol.* 1990; 294:607–622. [PubMed: 2341628]
- Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J. Neurosci.* 2000; 20:2369–2382. [PubMed: 10704511]
- Hart AS, Rutledge RB, Glimcher PW, Phillips PE. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J. Neurosci.* 2014; 34:698–704. [PubMed: 24431428]
- Heather N. A conceptual framework for explaining drug addiction. *J Psychopharmacol.* 1998; 12:3–7. [PubMed: 9584962]
- Heien ML, Khan AS, Ariansen JL, Cheer JF, Phillips PE, Wassum KM, Wightman RM. Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:10023–10028. [PubMed: 16006505]
- Henny P, Brown MT, Northrop A, Faunes M, Ungless MA, Magill PJ, Bolam JP. Structural correlates of heterogeneous in vivo activity of midbrain dopaminergic neurons. *Nat. Neurosci.* 2012; 15:613–619. [PubMed: 22327472]
- Holland PC. Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *Q. J. Exp. Psychol. B.* 1999; 52:307–333. [PubMed: 10605392]
- Hong S, Zhou TC, Smith M, Saleem KS, Hikosaka O. Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *J. Neurosci.* 2011; 31:11457–11471. [PubMed: 21832176]
- Hong S, Hikosaka O. Pedunculo-pontine tegmental nucleus neurons provide reward, sensorimotor, and alerting signals to midbrain dopamine neurons. *Neuroscience.* 2014; 282C:139–155. [PubMed: 25058502]
- Horger BA, Giles MK, Schenk S. Preexposure to amphetamine and nicotine predisposes rats to self-administer a low dose of cocaine. *Psychopharmacology (Berl).* 1992; 107:271–276. [PubMed: 1615126]
- Horger BA, Shelton K, Schenk S. Preexposure sensitizes rats to the rewarding effects of cocaine. *Pharmacol. Biochem. Behav.* 1990; 37:707–711. [PubMed: 2093176]
- Humphries MD, Prescott TJ. The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Prog. Neurobiol.* 2010; 90:385–417. [PubMed: 19941931]
- Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: The role of reward-related learning and memory. *Annu. Rev. Neurosci.* 2006; 29:565–598. [PubMed: 16776597]
- Ikemoto S. Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res. Rev.* 2007; 56:27–78. [PubMed: 17574681]
- Ilango A, Kesner AJ, Keller KL, Stuber GD, Bonci A, Ikemoto S. Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J. Neurosci.* 2014; 34:817–822. [PubMed: 24431440]
- Ito R, Dalley JW, Howes SR, Robbins TW, Everitt BJ. Dissociation in conditioned dopamine release in the nucleus accumbens core and shell in response to cocaine cues and during cocaine-seeking behavior in rats. *J. Neurosci.* 2000; 20:7489–7495. [PubMed: 11007908]
- Ito R, Dalley JW, Robbins TW, Everitt BJ. Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. *J. Neurosci.* 2002; 22:6247–6253. [PubMed: 12122083]

- Jennings JH, Ung RL, Resendez SL, Stamatakis AM, Taylor JG, Huang J, Veleta K, Katak PA, Aita M, Shilling-Scriver K, et al. Visualizing hypothalamic network dynamics for appetitive and consummatory behaviors. *Cell*. 2015; 160:516–527. [PubMed: 25635459]
- Jhou TC, Fields HL, Baxter MG, Saper CB, Holland PC. The rostromedial tegmental nucleus (rMTG), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron*. 2009; 61:786–800. [PubMed: 19285474]
- Kalivas PW. The glutamate homeostasis hypothesis of addiction. *Nat Rev Neurosci*. 2009; 10:561–572. [PubMed: 19571793]
- Kalivas PW, Duffy P. Effect of acute and daily cocaine treatment on extracellular dopamine in the nucleus accumbens. *Synapse*. 1990; 5:48–58. [PubMed: 2300906]
- Kamin, LJ. Predictability, surprise, attention, and conditioning. In: Cambell, BA.; Church, RM., editors. *Punishment and aversive behavior*. New York: Appleton-Century-Crofts; 1969. p. 279-296.
- Kempadoo KA, Tourino C, Cho SL, Magnani F, Leininger G- Stuber GD, Zhang F, Myers MG, Deisseroth K, Lecea L, Bonci A. Hypothalamic neurotensin projections promote reward by enhancing glutamate transmission in the VTA. *J. Neurosci*. 2013; 33:7618–7626. [PubMed: 23637156]
- Keramati M, Gutkin B. Imbalanced decision hierarchy in addicts emerging from drug-hijacked dopamine spiraling circuit. *PLoS ONE*. 2013; 8:e61489. [PubMed: 23637842]
- Kitai ST, Shepard PD, Callaway JC, Scroggs R. Afferent modulation of dopamine neuron firing patterns. *Curr. Opin. Neurobiol*. 1999; 9:690–697. [PubMed: 10607649]
- Kobayashi S, Schultz W. Reward contexts extend dopamine signals to unrewarded stimuli. *Curr. Biol*. 2014; 24:56–62. [PubMed: 24332545]
- Kobayashi Y, Okada K. Reward prediction error computation in the pedunculo-pontine tegmental nucleus neurons. *Ann. N. Y. Acad. Sci*. 2007; 1104:310–323. [PubMed: 17344541]
- Koob GF, Le Moal M. Neurobiological mechanisms for opponent motivational processes in addiction. *Philos Trans R Soc Lond B Biol Sci*. 2008; 363(1507):3113–3123. [PubMed: 18653439]
- Koob GF, Volkow ND. Neurocircuitry of addiction. *Neuropsychopharmacology*. 2010; 35:217–238. [PubMed: 19710631]
- Koulchitsky S, De Backer B, Quertemont E, Charlier C, Seutin V. Differential effects of cocaine on dopamine neuron firing in awake and anesthetized rats. *Neuropsychopharmacology*. 2012; 37:1559–1571. [PubMed: 22298123]
- Kravitz AV, Tye LD, Kreitzer AC. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat. Neurosci*. 2012; 15:816–818. [PubMed: 22544310]
- Kreitzer AC, Malenka RC. Striatal plasticity and basal ganglia circuit function. *Neuron*. 2008; 60:543–554. [PubMed: 19038213]
- Lammel S, Ion DI, Roeper J, Malenka RC. Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron*. 2011; 70:855–862. [PubMed: 21658580]
- Lammel S, Lim BK, Ran C, Huang KW, Betley MJ, Tye KM, Deisseroth K, Malenka RC. Input-specific control of reward and aversion in the ventral tegmental area. *Nature*. 2012; 491:212–217. [PubMed: 23064228]
- Lashley, KS. The problem of serial order in behavior. In: Jeffress, LA., editor. *Cerebral mechanisms in behavior*. New York: Wiley; 1951. p. 112-131.
- LeBlanc KH, Maidment NT, Ostlund SB. Repeated cocaine exposure facilitates the expression of incentive motivation and induces habitual control in rats. *PLoS ONE*. 2013; 8:e61355. [PubMed: 23646106]
- Lee BR, Ma YY, Huang YH, Wang X, Otaka M, Ishikawa M, Neumann PA, Graziane NM, Brown TE, Suska A, et al. Maturation of silent synapses in amygdala-accumbens projection contributes to incubation of cocaine craving. *Nat. Neurosci*. 2013; 16:1644–1651. [PubMed: 24077564]
- Lee T, Cai LX, Lelyveld VS, Hai A, Jasanoff A. Molecular-level functional magnetic resonance imaging of dopaminergic signaling. *Science*. 2014; 344:533–535. [PubMed: 24786083]
- Lenoir M, Serre F, Cantin L, Ahmed SH. Intense sweetness surpasses cocaine reward. *PLoS ONE*. 2007; 2:e698. [PubMed: 17668074]

- Liu QS, Pu L, Poo MM. Repeated cocaine exposure in vivo facilitates LTP induction in midbrain dopamine neurons. *Nature*. 2005; 437:1027–1031. [PubMed: 16222299]
- Ljungberg T, Apicella P, Schultz W. Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.* 1992; 67:145–163. [PubMed: 1552316]
- Lovinger DM. Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. *Neuropharmacology*. 2010; 58:951–961. [PubMed: 20096294]
- Luscher C. Drug-evoked synaptic plasticity causing addictive behavior. *J. Neurosci.* 2013; 33:17641–17646. [PubMed: 24198356]
- Ma YY, Lee BR, Wang X, Guo C, Liu L, Cui R, Lan Y, Balcita-Pedicino JJ, Wolf ME, Sesack SR, et al. Bidirectional modulation of incubation of cocaine craving by silent synapse-based remodeling of prefrontal cortex to accumbens projections. *Neuron*. 2014; 83:1453–1467. [PubMed: 25199705]
- Mahler SV, Vazey EM, Beckley JT, Keistler CR, McGlinchey EM, Kaufling J, Wilson SP, Deisseroth K, Woodward JJ, Aston-Jones G. Designer receptors show role for ventral pallidum input to ventral tegmental area in cocaine seeking. *Nat. Neurosci.* 2014; 17:577–585. [PubMed: 24584054]
- Mameli M, Halbout B, Creton C, Engblom D, Parkitna JR, Spanagel R, Luscher C. Cocaine-evoked synaptic plasticity: Persistence in the VTA triggers adaptations in the NAC. *Nat. Neurosci.* 2009; 12:1036–1041. [PubMed: 19597494]
- Mameli M, Luscher C. Synaptic plasticity and addiction: Learning mechanisms gone awry. *Neuropharmacology*. 2011; 61:1052–1059. [PubMed: 21277315]
- Matsumoto M, Hikosaka O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*. 2007; 447:1111–1115. [PubMed: 17522629]
- Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*. 2009; 459:837–841. [PubMed: 19448610]
- McClure SM, Daw ND, Montague PR. A computational substrate for incentive salience. *Trends Neurosci.* 2003; 26:423–428. [PubMed: 12900173]
- McCutcheon JE, Cone JJ, Sinon CG, Fortin SM, Kantak PA, Witten IB, Deisseroth K, Stuber GD, Roitman MF. Optical suppression of drug-evoked phasic dopamine release. *Front. Neural Circuits*. 2014; 8:114. [PubMed: 25278845]
- McNally GP, Johansen JP, Blair HT. Placing prediction into the fear circuit. *Trends Neurosci.* 2011; 34:283–292. [PubMed: 21549434]
- Miller RR, Barnet RC, Grahame NJ. Assessment of the Rescorla-Wagner model. *Psychol. Bull.* 1995; 117:363–386. [PubMed: 7777644]
- Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 1996; 16:1936–1947. [PubMed: 8774460]
- Morita K, Morishima M, Sakai K, Kawaguchi Y. Reinforcement learning: Computing the temporal difference of values via distinct corticostriatal pathways. *Trends Neurosci.* 2012; 35:457–467. [PubMed: 22658226]
- Morita K, Morishima M, Sakai K, Kawaguchi Y. Dopaminergic control of motivation and reinforcement learning: A closed-circuit account for reward-oriented behavior. *J. Neurosci.* 2013; 33:8866–8890. [PubMed: 23678129]
- Muscat R, Iravani MM, Kruk ZL. The role of ventral tegmental dopamine neurons in locomotor sensitization following quinpirole or (+)-amphetamine: Ex vivo voltammetric evidence. *Neuroscience*. 1996; 75:1175–1184. [PubMed: 8938750]
- Nair-Roberts RG, Chatelain-Badie SD, Benson E, White-Cooper H, Bolam JP, Ungless MA. Stereological estimates of dopaminergic, GABAergic and glutamatergic neurons in the ventral tegmental area, substantia nigra and retrorubral field in the rat. *Neuroscience*. 2008; 152:1024–1031. [PubMed: 18355970]
- Nakamura K, Ono T. Lateral hypothalamus neuron involvement in integration of natural and artificial rewards and cue signals. *J. Neurophysiol.* 1986; 55:163–181. [PubMed: 3512788]
- Nelson A, Killcross S. Amphetamine exposure enhances habit formation. *J. Neurosci.* 2006; 26:3805–3812. [PubMed: 16597734]
- Nestler EJ. Is there a common molecular pathway for addiction? *Nat. Neurosci.* 2005; 8:1445–1449. [PubMed: 16251986]

- Ng JP, Hubert GW, Justice JB Jr. Increased stimulated release and uptake of dopamine in nucleus accumbens after repeated cocaine administration as measured by in vivo voltammetry. *J. Neurochem.* 1991; 56:1485–1492. [PubMed: 2013751]
- Nicola SM, Surmeier J, Malenka RC. Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annu. Rev. Neurosci.* 2000; 23:185–215. [PubMed: 10845063]
- Nicola SM, Yun IA, Wakabayashi KT, Fields HL. Cue-evoked firing of nucleus accumbens neurons encodes motivational significance during a discriminative stimulus task. *J. Neurophysiol.* 2004; 91:1840–1865. [PubMed: 14645377]
- Nieh EH, Matthews GA, Allsop SA, Presbrey KN, Leppla CA, Wichmann R, Neve R, Wildes CP, Tye KM. Decoding neural circuits that control compulsive sucrose seeking. *Cell.* 2015; 160:528–541. [PubMed: 25635460]
- Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology (Berl).* 2007; 191:507–520. [PubMed: 17031711]
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science.* 2004; 304:452–454. [PubMed: 15087550]
- Olausson P, Jentsch JD, Tronson N, Neve RL, Nestler EJ, Taylor JR. Deltafosb in the nucleus accumbens regulates food-reinforced instrumental behavior and motivation. *J. Neurosci.* 2006; 26:9196–9204. [PubMed: 16957076]
- Pan B, Hillard CJ, Liu QS. Endocannabinoid signaling mediates cocaine-induced inhibitory synaptic plasticity in midbrain dopamine neurons. *J. Neurosci.* 2008; 28:1385–1397. [PubMed: 18256258]
- Pan WX, Hyland BI. Pedunculopontine tegmental nucleus controls conditioned responses of midbrain dopamine neurons in behaving rats. *J. Neurosci.* 2005; 25:4725–4732. [PubMed: 15888648]
- Pan WX, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: Evidence for eligibility traces in the reward-learning network. *J. Neurosci.* 2005; 25:6235–6242. [PubMed: 15987953]
- Pascoli V, Terrier J, Espallergues J, Valjent E, O'Connor EC, Luscher C. Contrasting forms of cocaine-evoked plasticity control components of relapse. *Nature.* 2014; 509:459–464. [PubMed: 24848058]
- Perry AN, Westenbroek C, Becker JB. The development of a preference for cocaine over food identifies individual rats with addiction-like behaviors. *PLoS ONE.* 2013; 8:e79465. [PubMed: 24260227]
- Phillips PE, Stuber GD, Heien ML, Wightman RM, Carelli RM. Subsecond dopamine release promotes cocaine seeking. *Nature.* 2003; 422:614–618. [PubMed: 12687000]
- Piazza PV, Deminiere JM, Le Moal M, Simon H. Factors that predict individual vulnerability to amphetamine self-administration. *Science.* 1989; 245:1511–1513. [PubMed: 2781295]
- Pierce RC, Kumaresan V. The mesolimbic dopamine system: The final common pathway for the reinforcing effect of drugs of abuse? *Neurosci Biobehav Rev.* 2006; 30:215–238. [PubMed: 16099045]
- Pignatelli M, Bonci A. Role of Dopamine Neurons in Reward and Aversion: A Synaptic Plasticity Perspective. *Neuron.* 2015; 86(5):1145–1157. [PubMed: 26050034]
- Porrino L, Lyons D, Smith H, Daunais J, Nader M. Cocaine self-administration produces a progressive involvement of limbic, association, and sensorimotor striatal domains. *J. Neurosci.* 2004; 24:3554–3562. [PubMed: 15071103]
- Redish AD. Addiction as a computational process gone awry. *Science.* 2004; 306:1944–1947. [PubMed: 15591205]
- Rescorla, RA.; Wagner, AR. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH.; Prokasy, WF., editors. *Classical conditioning ii: Current research and theory.* New York: Appleton Century Crofts; 1972. p. 64-99.
- Robbins TW, Everitt BJ. Functions of dopamine in the dorsal and ventral striatum. *Seminars in Neuroscience.* 1992; 4:119–127.
- Robbins TW, Everitt BJ. A role for mesencephalic dopamine in activation: Commentary on berridge. *Psychopharmacology (Berl).* 2006; 191:433–437. [PubMed: 16977476]

- Robinson TE, Becker JB. Enduring changes in brain and behavior produced by chronic amphetamine administration: A review and evaluation of animal models of amphetamine psychosis. *Brain Res.* 1986; 396:157–198. [PubMed: 3527341]
- Robinson TE, Berridge KC. Incentive-sensitization and addiction. *Addiction.* 2001; 96:103–114. [PubMed: 11177523]
- Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G. Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *J. Neurosci.* 2009; 29:13365–13376. [PubMed: 19846724]
- Roitman MF, Stuber GD, Phillips PE, Wightman RM, Carelli RM. Dopamine operates as a subsecond modulator of food seeking. *J. Neurosci.* 2004; 24:1265–1271. [PubMed: 14960596]
- Rolls ET, Burton MJ, Mora F. Hypothalamic neuronal responses associated with the sight of food. *Brain Res.* 1976; 111:53–66. [PubMed: 821581]
- Saal D, Dong Y, Bonci A, Malenka RC. Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. *Neuron.* 2003; 37:577–582. [PubMed: 12597856]
- Salamone J, Correa M. The mysterious motivational functions of mesolimbic dopamine. *Neuron.* 2012; 76:470–485. [PubMed: 23141060]
- Schilstrom B, Yaka R, Argilli E, Suvarna N, Schumann J, Chen BT, Carman M, Singh V, Mailliard WS, Ron D, Bonci A. Cocaine enhances NMDA receptor-mediated currents in ventral tegmental area cells via dopamine D5 receptor-dependent redistribution of NMDA receptors. *J. Neurosci.* 2006; 26:8549–8558. [PubMed: 16914681]
- Schoenbaum G, Setlow B. Cocaine makes actions insensitive to outcomes but not extinction: Implications for altered orbitofrontal-amygdalar function. *Cereb. Cortex.* 2005; 15:1162–1169. [PubMed: 15563719]
- Schultz W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 1998; 80:1–27. [PubMed: 9658025]
- Schultz W. Behavioral dopamine signals. *Trends Neurosci.* 2007; 30:203–210. [PubMed: 17400301]
- Schultz W. Updating dopamine reward signals. *Curr. Opin. Neurobiol.* 2013; 23:229–238. [PubMed: 23267662]
- Schultz W, Apicella P, Ljungberg T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* 1993; 13:900–913. [PubMed: 8441015]
- Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science.* 1997; 275:1593–1599. [PubMed: 9054347]
- Schultz W, Dickinson A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 2000; 23:473–500. [PubMed: 10845072]
- Sesack SR, Grace AA. Cortico-basal ganglia reward network: Microcircuitry. *Neuropsychopharmacology.* 2010; 35:27–47. [PubMed: 19675534]
- Shen HW, Gipson CD, Huits M, Kalivas PW. Prelimbic cortex and ventral tegmental area modulate synaptic plasticity differentially in nucleus accumbens during cocaine-reinstated drug seeking. *Neuropsychopharmacology.* 2014; 39:1169–1177. [PubMed: 24232172]
- Skog, O-J. "Addiction: Definition and mechanisms". In: Heather, N.; Vuchinich, RE., editors. *Choice, behavioural economics and addiction.* Oxford: Elsevier; 2003. p. 157-175.
- Somers LA, Beyene M, Carelli RM, Wightman RM. Synaptic overflow of dopamine in the nucleus accumbens arises from neuronal activity in the ventral tegmental area. *J. Neurosci.* 2009; 29:1735–1742. [PubMed: 19211880]
- Stacy AW, Wiers RW. Implicit cognition and addiction: A tool for explaining paradoxical behavior. *Annu. Rev. Clin. Psychol.* 2010; 6:551–575. [PubMed: 20192786]
- Stalnaker TA, Calhoun GG, Ogawa M, Roesch MR, Schoenbaum G. Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Front. Integr. Neurosci.* 2010; 4:12. [PubMed: 20508747]
- Steinberg EE, Boivin JR, Saunders BT, Witten IB, Deisseroth K, Janak PH. Positive reinforcement mediated by midbrain dopamine neurons requires D1 and D2 receptor activation in the nucleus accumbens. *PLoS ONE.* 2014; 9:e94771. [PubMed: 24733061]

- Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* 2013; 16:966–973. [PubMed: 23708143]
- Stuber GD, Roitman MF, Phillips PE, Carelli RM, Wightman RM. Rapid dopamine signaling in the nucleus accumbens during contingent and noncontingent cocaine administration. *Neuropsychopharmacology.* 2005a; 30:853–863. [PubMed: 15549053]
- Stuber GD, Wightman RM, Carelli RM. Extinction of cocaine self-administration reveals functionally and temporally distinct dopaminergic signals in the nucleus accumbens. *Neuron.* 2005b; 46:661–669. [PubMed: 15944133]
- Sulzer D. How addictive drugs disrupt presynaptic dopamine neurotransmission. *Neuron.* 2011; 69:628–649. [PubMed: 21338876]
- Sutton, R.; Barto, A. Reinforcement learning: An introduction. MIT Press; 1998.
- Swanson LW. The projections of the ventral tegmental area and adjacent regions: A combined fluorescent retrograde tracer and immunofluorescence study in the rat. *Brain Res. Bull.* 1982; 9:321–353. [PubMed: 6816390]
- Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat. Neurosci.* 2011; 14:1590–1597. [PubMed: 22037501]
- Takahashi Y, Geoffrey Schoenbaum G, Niv Y. Silencing the Critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an Actor/Critic model *Front. Neurosci.* 2008; 2(1):86–99.
- Tan K, Yvon C, Turiault M, Mirzabekov J, Doehner J, Labouèbe G, Deisseroth K, Tye K, Lüscher C. GABA neurons of the VTA drive conditioned place aversion. *Neuron.* 2012; 73:1173–1183. [PubMed: 22445344]
- Taylor JR, Jentsch JD. Repeated intermittent administration of psychomotor stimulant drugs alters the acquisition of pavlovian approach behavior in rats: Differential effects of cocaine, d-amphetamine and 3,4- methylenedioxymethamphetamine ("ecstasy"). *Biol. Psychiatry.* 2001; 50:137–143. [PubMed: 11526995]
- Tepper JM, Lee CR. GABAergic control of substantia nigra dopaminergic neurons. *Prog. Brain Res.* 2007; 160:189–208. [PubMed: 17499115]
- Threlfell S, Cragg SJ. Dopamine signaling in dorsal versus ventral striatum: The dynamic role of cholinergic interneurons. *Front. Syst. Neurosci.* 2011; 5:11. [PubMed: 21427783]
- Tian J, Uchida N. Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. *Neuron.* in press
- Timberlake W. Behavior systems, associationism, and pavlovian conditioning. *Psychon. Bull. Rev.* 1994; 1:405–420. [PubMed: 24203549]
- Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science.* 2009; 324:1080–1084. [PubMed: 19389999]
- Ungless MA, Whistler JL, Malenka RC, Bonci A. Single cocaine exposure in vivo induces long-term potentiation in dopamine neurons. *Nature.* 2001; 411:583–587. [PubMed: 11385572]
- Vander Weele CM, Porter-Stransky KA, Mabrouk OS, Lovic V, Singer BF, Kennedy RT, Aragona BJ. Rapid dopamine transmission within the nucleus accumbens: dramatic difference between morphine and oxycodone delivery. *Eur J Neurosci.* 2014; 40(7):3041–3054. [PubMed: 25208732]
- van Zessen R, Phillips JL, Budygin EA, Stuber GD. Activation of VTA GABA neurons disrupts reward consumption. *Neuron.* 2012; 73:1184–1194. [PubMed: 22445345]
- Vanderschuren LJ, Everitt BJ. Drug seeking becomes compulsive after prolonged cocaine self-administration. *Science.* 2004; 305:1017–1019. [PubMed: 15310907]
- Vanderschuren LJ, Kalivas PW. Alterations in dopaminergic and glutamatergic transmission in the induction and expression of behavioral sensitization: A critical review of preclinical studies. *Psychopharmacology (Berl).* 2000; 151:99–120. [PubMed: 10972458]

- Vezina P, Pierre PJ, Lorrain DS. The effect of previous exposure to amphetamine on drug-induced locomotion and self-administration of a low dose of the drug. *Psychopharmacology (Berl)*. 1999; 147:125–134. [PubMed: 10591879]
- Volkow ND, Wang GJ, Telang F, Fowler JS, Logan J, Childress AR, Jayne M, Ma Y, Wong C. Cocaine cues and dopamine in dorsal striatum: Mechanism of craving in cocaine addiction. *J. Neurosci*. 2006; 26:6583–6588. [PubMed: 16775146]
- Vollstadt-Klein S, Wichert S, Rabinstein J, Buhler M, Klein O, Ende G, Hermann D, Mann K. Initial, habitual and compulsive alcohol use is characterized by a shift of cue processing from ventral to dorsal striatum. *Addiction*. 2010; 105:1741–1749. [PubMed: 20670348]
- Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature*. 2001; 412:43–48. [PubMed: 11452299]
- Walker DM, Cates HM, Heller EA, Nestler EJ. Regulation of chromatin states by drugs of abuse. *Curr. Opin. Neurobiol*. 2015; 30:112–121. [PubMed: 25486626]
- Wanat MJ, Willuhn I, Clark JJ, Phillips PE. Phasic dopamine release in appetitive behaviors and drug addiction. *Curr. Drug Abuse Rev*. 2009; 2:195–213. [PubMed: 19630749]
- Wang LP, Li F, Wang D, Xie K, Wang D, Shen X, Tsien JZ. NMDA receptors in dopaminergic neurons are crucial for habit learning. *Neuron*. 2011; 72:1055–1066. [PubMed: 22196339]
- Wassum KM, Ostlund SB, Maidment NT. Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. *Biol. Psychiatry*. 2012; 71:846–854. [PubMed: 22305286]
- Watabe-Uchida M, Zhu L, Ogawa S, Vamanrao A, Uchida N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron*. 2012; 74:858–873. [PubMed: 22681690]
- White FJ, Hu XT, Zhang XF, Wolf ME. Repeated administration of cocaine or amphetamine alters neuronal responses to glutamate in the mesoaccumbens dopamine system. *J. Pharmacol. Exp. Ther*. 1995; 273:445–454. [PubMed: 7714800]
- Wickens JR. Synaptic plasticity in the basal ganglia. *Behav. Brain Res*. 2009; 199:119–128. [PubMed: 19026691]
- Williams JE, Wiczorek W, Willner P, Kruk ZL. Parametric analysis of the effects of cocaine and cocaine pretreatment on dopamine release in the nucleus accumbens measured by fast cyclic voltammetry. *Brain Res*. 1995; 678:225–232. [PubMed: 7620892]
- Willuhn I, Burgeno LM, Everitt BJ, Phillips PE. Hierarchical recruitment of phasic dopamine signaling in the striatum during the progression of cocaine use. *Proc. Nat. Acad. Sci. U.S.A*. 2012; 109:20703–20708.
- Willuhn I, Burgeno LM, Groblewski PA, Phillips PE. Excessive cocaine use results from decreased phasic dopamine signaling in the striatum. *Nat. Neurosci*. 2014; 17:704–709. [PubMed: 24705184]
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron*. 2014; 81:267–279. [PubMed: 24462094]
- Wise RA. Dopamine, learning and motivation. *Nat Rev Neurosci*. 2004; 5:483–494. [PubMed: 15152198]
- Witten IB, Steinberg EE, Lee SY, Davidson TJ, Zalocusky KA, Brodsky M, Yizhar O, Cho SL, Gong S, Ramakrishnan C, et al. Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron*. 2011; 72:721–733. [PubMed: 22153370]
- Wong DF, Kuwabara H, Schretlen DJ, Bonson KR, Zhou Y, Nandi A, Brasic JR, Kimes AS, Maris MA, Kumar A, et al. Increased occupancy of dopamine receptors in human striatum during cue-elicited cocaine craving. *Neuropsychopharmacology*. 2006; 31:2716–2727. [PubMed: 16971900]
- Xia Y, Driscoll JR, Wilbrecht L, Margolis EB, Fields HL, Hjelmstad GO. Nucleus accumbens medium spiny neurons target non-dopaminergic neurons in the ventral tegmental area. *J. Neurosci*. 2011; 31:7811–7816. [PubMed: 21613494]
- Yamaguchi T, Qi J, Wang HL, Zhang S, Morales M. Glutamatergic and dopaminergic neurons in the mouse ventral tegmental area. *Eur. J. Neurosci*. 2015; 41:760–772. [PubMed: 25572002]
- Yamaguchi T, Wang H, Morales M. Glutamate neurons in the substantia nigra compacta and retrorubral field. *Eur. J. Neurosci*. 2013; 38:3602–3610. [PubMed: 24102658]

- Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 2006; 7:464–476. [PubMed: 16715055]
- Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 2004; 19:181–189. [PubMed: 14750976]
- Zapata A, Minney VL, Shippenberg TS. Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. *J. Neurosci.* 2010; 30:15457–15463. [PubMed: 21084602]
- Zernig G, Ahmed SH, Cardinal RN, Morgan D, Acquas E, Foltin RW, Vezina P, Negus SS, Crespo JA, Stockl P, et al. Explaining the escalation of drug use in substance dependence: Models and appropriate animal laboratory tests. *Pharmacology.* 2007; 80:65–119. [PubMed: 17570954]
- Zweifel LS, Parker JG, Lobb CJ, Rainwater A, Wall ZV, Fadok JP, Darvas M, Kim MJ, Mizumori SJ, Paladini CA, et al. Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc. Nat. Acad. Sci. U.S.A.* 2009; 106:7281–7288.

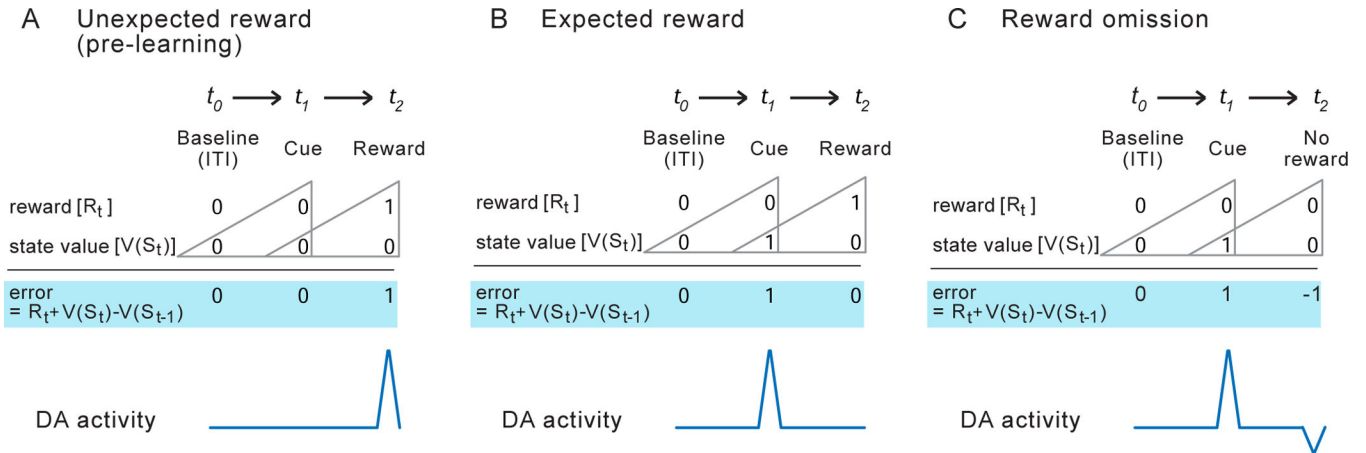


Figure 1. Activity of DA neurons complies with formal models of reinforcement

The temporal difference model of reinforcement defines a reward prediction error (RPE) as the discrepancy between the most recent reward prediction $[V(S_{t-1})]$ and any new information regarding reward, be it the reward itself $[R_t]$ or a signal that causes a change in the prospect of reward $[V(S_t)]$. **A.** Early in Pavlovian training, the surprising delivery of reward produces a positive RPE paralleled by phasic activation of DA neurons. **B.** With sufficient training the presentation of a cue signaling reward produces a positive RPE while the reward itself no longer results in RPE. This shift is paralleled by a similar shift in the DA neuron response. **C.** Omission of expected reward results in a negative RPE, paralleled by a transient reduction of DA activity below baseline. ITI= Inter Trial Interval. (Adapted from Morita et al. 2012.)

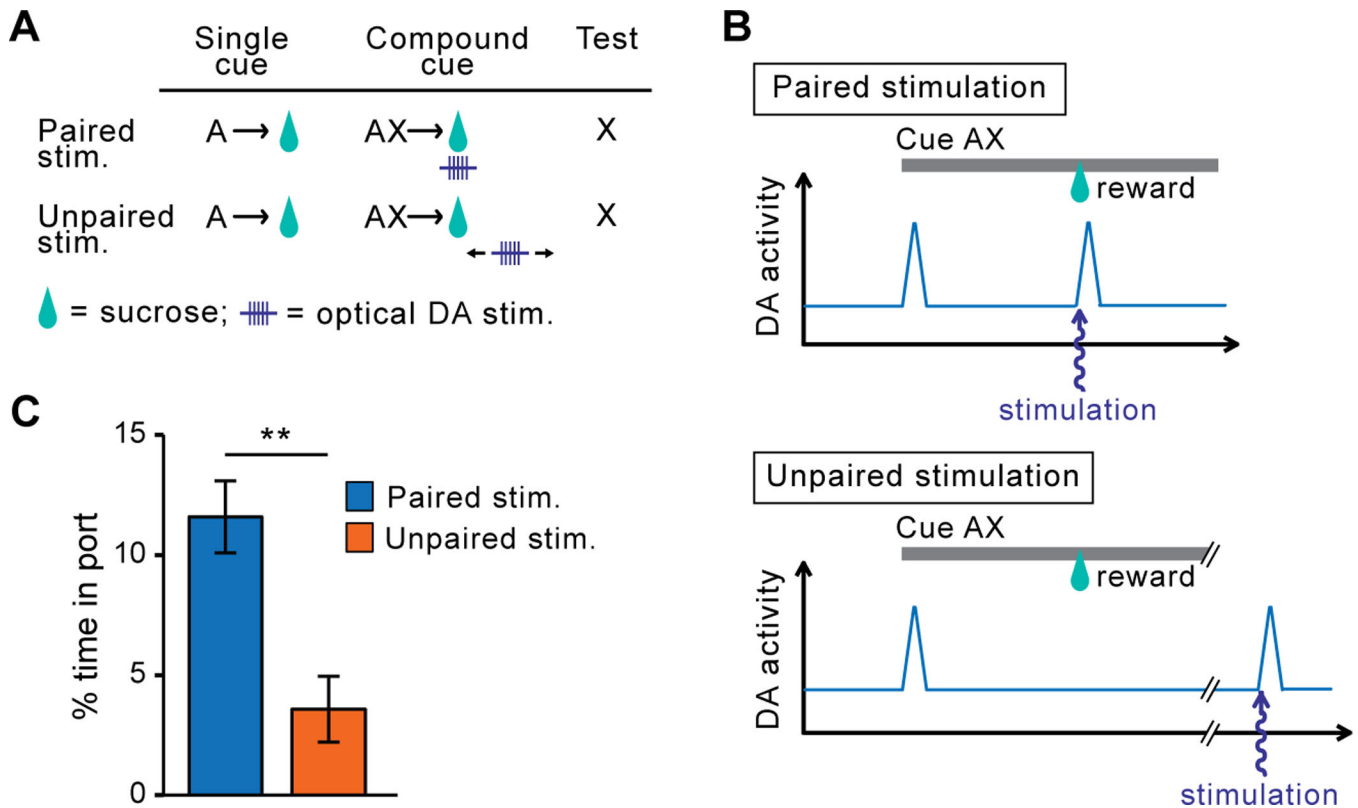
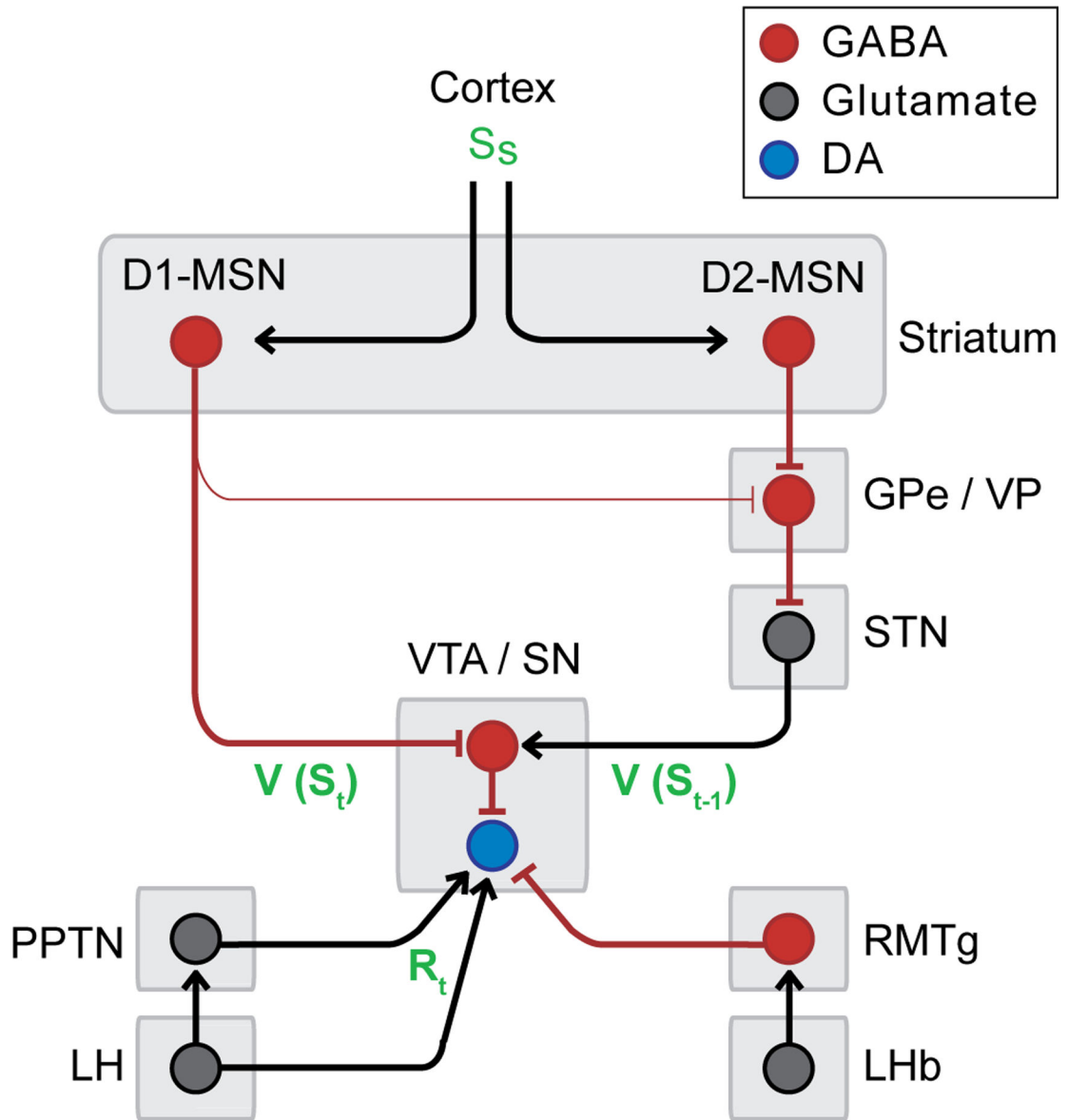


Figure 2. Optogenetic stimulation of DA neurons mimics RPE and drives reward learning

A. Behavioral protocol. Rats were trained to associate an auditory stimulus (cue A) with sucrose. Once this association was learned, as attested by stable levels of conditioned approach to the sucrose delivery port, a visual stimulus (cue X) was added to the auditory stimulus and this compound stimulus (cue AX) was paired with the sucrose reward. During compound stimulus conditioning sessions, DA neurons were photoactivated during reward consumption (Paired Stim.) to artificially create a normally-absent RPE, the sucrose reward being perfectly predicted at this stage by cue A. Control (Unpaired Stim.) rats received DA neuron activation of DA neurons during the intertrial interval. At test, conditioned responding to stimulus X alone was tested in absence of sucrose or optical stimulation. **B.** Optical stimulation of DA neurons during a compound cue trial. Optical stimulation of DA neurons was synchronized with the delivery of the anticipated reward in Paired Stimulation rats; Unpaired Stimulation rats received optical stimulation of DA neurons at a variable time after cue and reward delivery. **C.** In rats that received previous stimulation of DA neurons during reward, the presentation of cue X elicited approach to the location of previous sucrose delivery. In contrast rats that received DA neuron manipulation unpaired with reward showed low, or blocked, responding to cue X (Steinberg et al., 2013).



$$\text{Prediction Error} = R_t + V(S_t) - V(S_{t-1})$$

Figure 3. Simplified neural circuit diagram for the computation of the RPE by DA neurons
 The coordinated activation of this circuit by rewards and their predictors could result in DA neuron firing in accordance with the RPE model. GPe: external globus pallidus; LHb: lateral habenula; LH: lateral hypothalamus; MSN: medium spiny neurons; PPTN: pedunculopontine tegmental nucleus; RMTg: rostromedial tegmental nucleus; SN: substantia nigra; STN: subthalamic nucleus; VP: ventral pallidum; VTA: ventral tegmental area

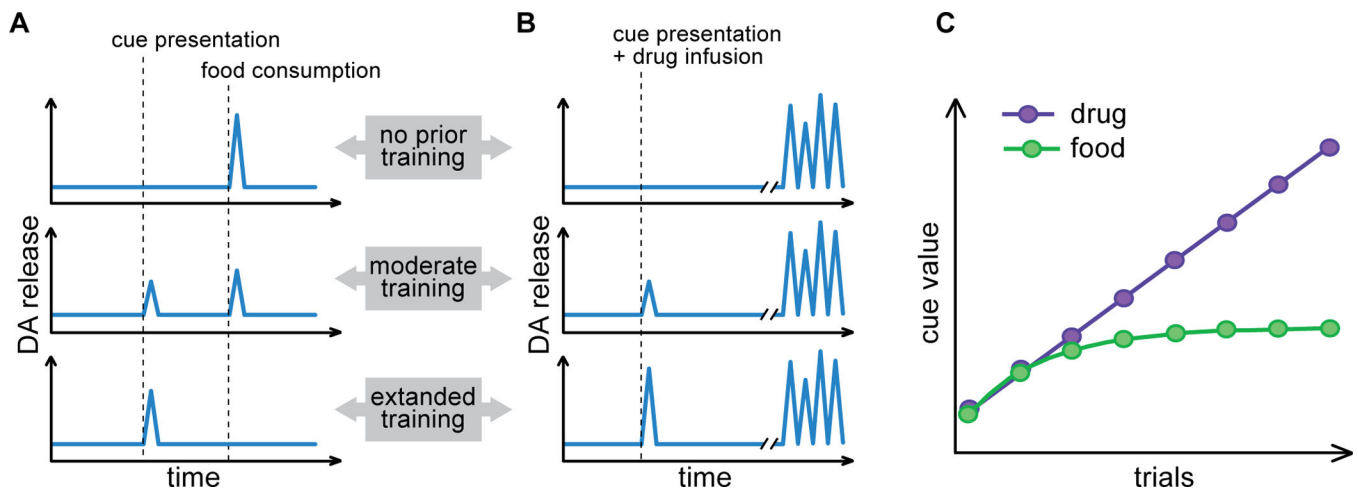


Figure 4. Differences between food-related and cocaine-related phasic DA signals in the NAc and potential consequences for learning

A. Prior to learning unexpected delivery of food reward results in phasic DA signals. As subjects learn that cue presentation signals food delivery, DA responses are transferred from the reward to the cue. **B.** When cocaine is the reward, each drug injection produces, with some delay, a burst of phasic DA events as a consequence of the pharmacological actions of the drug. As with natural reward, phasic DA responses progressively emerge to the cue. Unlike food-induced DA signals, drug-induced DA signals are not modulated by expectations and persist throughout learning. **C.** Proposed consequences of these DA signals on learning. Food-evoked DA signals modulated by reward expectations promote learning until the prediction matches the actual outcome, resulting in stable cue value after a few trials. In contrast, persistent cocaine-evoked DA signals continue to increase the value of cocaine cues with every trial. Eventually, the value of cocaine cues surpasses the value of the food cues and can bias decision-making towards cocaine.

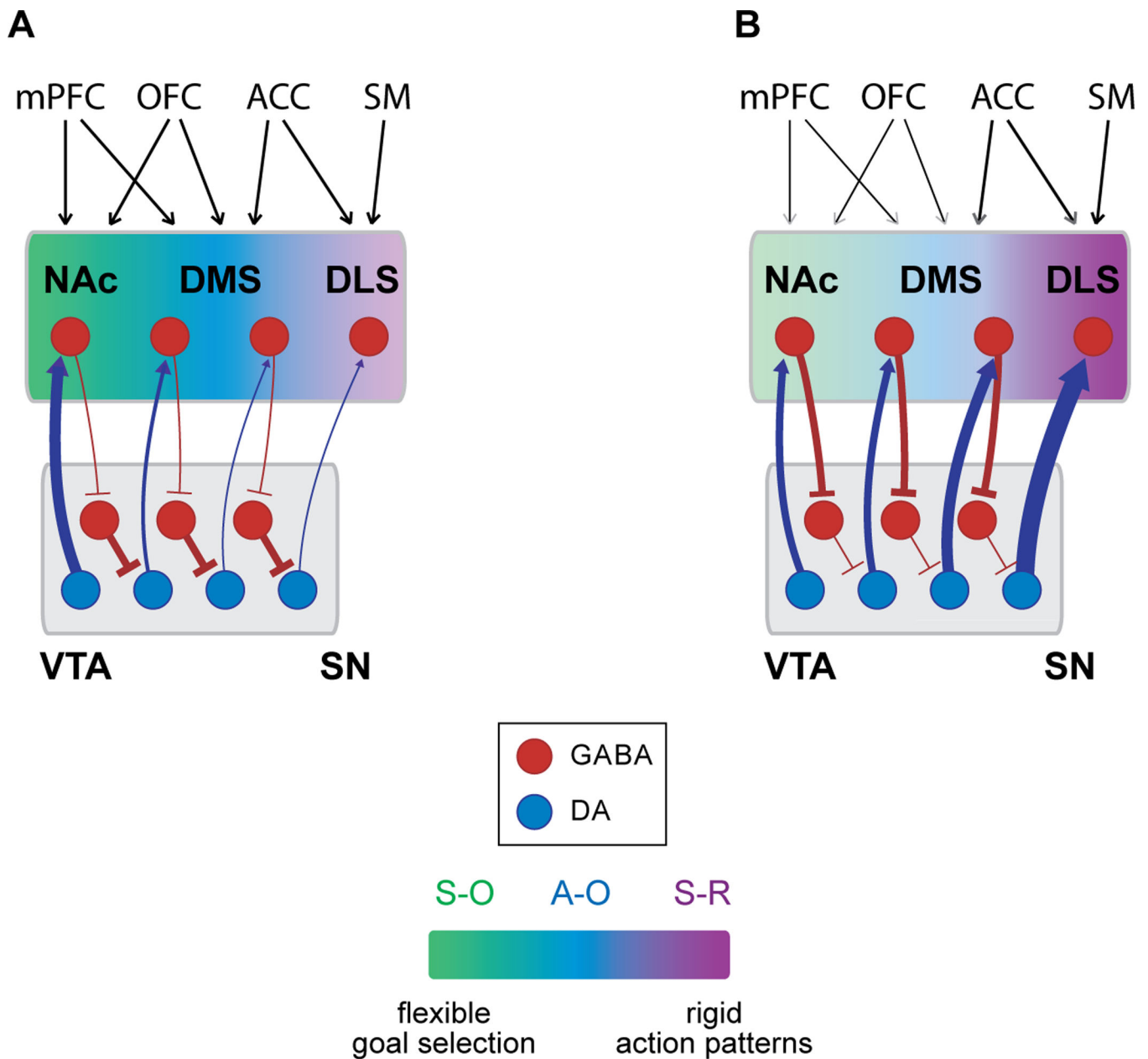


Figure 5. Proposed mechanism for the accelerated propagation of RPE-DA signals in different striatal domains following drug exposure and its consequence for learning

A. In drug naïve animals the activity of midbrain DA neurons is tightly regulated by local inhibitory GABA neurons. Unexpected rewards activate dopamine neurons in the VTA; the resulting DA release in the NAc promotes Pavlovian (S–O) learning. **B.** Repeated exposure to cocaine increases the excitability of DA neurons by potentiating striatal inhibitory inputs on midbrain local GABA neurons. Striatal feedback on DA neurons progressively recruits more lateral DA neurons from VTA to SN for encoding of RPE. The resulting emergence of DA signals in the sensorimotor DLS reinforces S–R associations that contribute to rigid and possibly compulsive drug-seeking behavior. ACC: anterior cingulate cortex; DLS: dorsolateral striatum; DMS: dorsomedial striatum; NAc: nucleus accumbens; OFC:

orbitofrontal cortex; SM: sensorymotor cortex; SN: substantia nigra; VTA: ventral tegmental area; S-O: stimulus-outcome; A-O: action-outcome; S-R: stimulus-response.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript