Neuroscience and Biobehavioral Reviews 34 (2010) 701–720

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Neural mechanisms of acquired phasic dopamine responses in learning☆

Thomas E. Hazy [a], Michael J. Frank [b,*], Randall C. O'Reilly [a]

[a] Department of Psychology and Neuroscience, University of Colorado Boulder, 345 UCB, Boulder, CO 80309, United States
[b] Departments of Psychology and Cognitive and Linguistic Sciences, Brown University, 190 Thayer Street, Providence, RI 02912, United States

ARTICLE INFO

Keywords:
Dopamine
Learning
Conditioning
Basal ganglia
Computational model

ABSTRACT

What biological mechanisms underlie the reward-predictive firing properties of midbrain dopaminergic neurons, and how do they relate to the complex constellation of empirical findings understood as Pavlovian and instrumental conditioning? We previously presented PVLV, a biologically inspired Pavlovian learning algorithm accounting for DA activity in terms of two interrelated systems: a primary value (PV) system, which governs how DA cells respond to a US (reward) and; a learned value (LV) system, which governs how DA cells respond to a CS. Here, we provide a more extensive review of the biological mechanisms supporting phasic DA firing and their relation to the spate of Pavlovian conditioning phenomena and their sensitivity to focal brain lesions. We further extend the model by incorporating a new NV (novelty value) component reflecting the ability of novel stimuli to trigger phasic DA firing, providing "novelty bonuses" which encourages exploratory working memory updating and in turn speeds learning in trace conditioning and other working memory-dependent paradigms. The evolving PVLV model builds upon insights developed in many earlier computational models, especially reinforcement learning models based on the ideas of Sutton and Barto, biological models, and the psychological model developed by Savastano and Miller. The PVLV framework synthesizes these various approaches, overcoming important shortcomings of each by providing a coherent and specific mapping to much of the relevant empirical data at both the micro- and macro-levels, and examines their relevance for higher order cognitive functions.

Contents

1. Introduction . . . 702
   1.1. Two critical questions—at two different levels of explanation . . . 702
   1.2. Organization of the paper . . . 703
2. Overview of the PVLV model . . . 704
   2.1. Biological interpretation of the PVLV model . . . 705
   2.2. Claim-by-claim review of PVLV-proposed mechanisms of phasic dopamine firing . . . 706
   2.3. PVLV hypothesis #1: PV_e representation ≈ LHA . . . 707
   2.4. PVLV hypothesis #2: PV_i representations ≈ patch-like neurons in the VS . . . 707
      2.4.1. Blocking phasic US bursting . . . 707
      2.4.2. Generating dips for omitted expected rewards . . . 708
   2.5. PVLV hypothesis #3: LV_e representation ≈ CNA . . . 709
   2.6. PVLV hypothesis #4: CNA multi-modal cells are not trained by CS-triggered DA . . . 709
   2.7. PVLV hypothesis #5: with overtraining, LV_i partially mitigates phasic DA firing to CS-onset . . . 710
   2.8. PVLV hypothesis #6: phasic DA bursting arises from sustained inputs . . . 710
3. The role of CS-onset phasic DA signals in conditioning and learning . . . 711
   3.1. The conditioned orienting response (COR) . . . 711
   3.2. Autoshaping . . . 712
   3.3. Training of CS-generated CRs through phasic DA . . . 712

☆ Supported by ONR grants N00014-07-1-0651 and N00014-03-1-0428, and NIH grants MH069597 and MH079485.
* Corresponding author.
E-mail addresses: michael_frank@brown.edu (M.J. Frank), Randy.OReilly@colorado.edu (R.C. O'Reilly).

0149-7634/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neubiorev.2009.11.019

3.4. CS-generated CRs are not susceptible to the blocking effect . . . 713
3.5. CNA lesions impair generalized PIT . . . 713
4. PVLV algorithm updates . . . 713
4.1. Novelty value (NV) . . . 713
4.2. $\dot{Y}$ mechanism for DA bursting . . . 714
4.3. Updated computational implementation . . . 714
5. Comparison between PVLV and TD . . . 715
6. Discussion . . . 716
6.1. Key implications and predictions of the PVLV framework . . . 716
6.2. Beyond the core PVLV model: ancillary mechanisms and future development . . . 717
6.3. Conclusion . . . 717
References . . . 717

1. Introduction

One of the seminal findings of the last decade and a half has been the discovery that midbrain dopamine (DA) neurons exhibit a pattern of firing that maps remarkably well to the *reward prediction error* signal first proposed by Rescorla and Wagner (1972) and since widely adopted as the dominant account of Pavlovian conditioning (e.g., Montague et al., 1997; Schultz et al., 1997; Schultz, 1998; Hollerman and Schultz, 1998; Schultz and Dickinson, 2000; Waelti et al., 2001). While phasic firing has also been described for other categories of salient stimuli besides positively-valenced reward (e.g., aversive, novel, or high intensity stimuli; Bayer and Glimcher, 2005; Bayer et al., 2007; Horvitz et al., 1997; Ljungberg et al., 1992; Steinfels et al., 1983; Legault and Wise, 2001; Horvitz, 2000; Pan et al., 2005; Satoh et al., 2003), the robust correlation with reward prediction error has generated a great deal of recent attention, in part due to its potential to explain both Pavlovian and instrumental conditioning phenomena, and the uncanny resemblance of phasic DA to signals developed in artificial reinforcement learning algorithms.

In addition, along with the recognition of the importance of DA signals for lower level phenomena like conditioning, DA signals have also been increasingly implicated in relatively higher level phenomena more often associated with cognitive neuroscience such as decision making, executive function and the new field of neuroeconomics (e.g., Braver and Cohen, 2000; McClure et al., 2005; Montague et al., 2004; Dayan and Niv, 2008; Glimcher, 2008). Thus, it seems likely that a better understanding of the mechanisms involved in DA signaling may help improve our understanding across multiple levels of explanation from basic physiology to complex cognitive behavior.

To summarize the now well known findings, when DA neurons are recorded in a Pavlovian paradigm, brief, phasic bursts of dopamine firing (above a tonic baseline level) are initially observed at the unconditioned stimulus (US; primary reward) onset. Over repeated trials, however, dopamine bursts are elicited at the onset of the conditioned stimulus (CS), while also showing attenuated responses to the US. By the end of training US-associated firing has disappeared completely, having effectively been transferred to the time of CS-onset. Fig. 1 summarizes these basic findings, based largely on data from Schultz and his group, (e.g., Schultz, 1998; Schultz et al., 1993a). Because of this behavior, many cases of phasic DA cell firing can be meaningfully interpreted as a *reward prediction error* signal, encoding the extent to which a given reward was unexpected at that particular time (e.g., Schultz, 1998; Hollerman and Schultz, 1998). Based on these results, in conjunction with dopamine's (and D1 receptor agonists') well documented ability to produce the late form of long term potentiation (L-LTP) at the synapse (e.g., Frey et al., 1990, 1991; Huang and Kandel, 1995; Wickens et al., 1996; Chong et al., 2006), there now exists a broad consensus that phasic DA firing is a learning signal for both Pavlovian and instrumental conditioning (e.g., Montague et al., 1997; Schultz, 1998; Waelti et al., 2001; Wickens et al., 2007). Critically, pharmacological DA manipulations robustly affect behavioral learning from positive and negative reward prediction errors (Frank et al., 2004; Cools et al., 2006; Frank and O'Reilly, 2006; Santesso et al., 2009; Moustafa et al., 2008; Bodi et al., 2009; Palminteri et al., 2009; Cools et al., 2009), implying that DA activations are more than just correlational.

1.1. Two critical questions—at two different levels of explanation

Despite the widespread acceptance of this hypothesis, however—a hypothesis that might be called the *phasic dopamine reward prediction error hypothesis of Pavlovian conditioning* —the mechanisms that may underly it remain inadequately characterized. Consider the following two questions, each at a different level of analysis:

1. What biological substrates and mechanisms provide the representations and projections that drive midbrain dopaminergic neurons to exhibit their reward-predictive firing properties? Importantly, we also need to know how that behavior can be modified over time in response to changing environmental contingencies (i.e., learning).
2. How do the resulting phasic DA signals actually contribute to the learning and behavioral effects so characteristic of Pavlovian (and instrumental) conditioning?

The first question treats the DA firing data as an *effect* and seeks its cause: "How might one account for this data based on lower-level mechanisms?" The second question treats phasic DA firing as a *cause* and seeks to understand its effects: "Taking the DA firing pattern as a given, how does it modulate function in downstream brain areas, and how can that account for phenomena at the higher, behavioral level?" Clearly, both levels of analysis are necessary for a full understanding.

With regard to the lower-level set of questions, the consensus among most researchers seems to be that the critical learning is taking place upstream from the midbrain dopamine neurons themselves—and several hypotheses have been put forward to explain some or all of the story (e.g., Brown et al., 1999; Houk et al., 1995; Miller, 2000; Schultz, 1998; Stuber et al., 2008; Tan and Bullock, 2008). Of these, the leading theoretical framework has been temporal differences (TD) algorithm of Sutton and Barto (Sutton, 1988; Sutton and Barto, 1990, 1998), which posits a unitary backward chaining mechanism of explicit timestep-to-timestep value predictions with a prediction error computed for each step. Although there is much evidence that DA neurons report signals that resemble reward prediction errors, there remains a

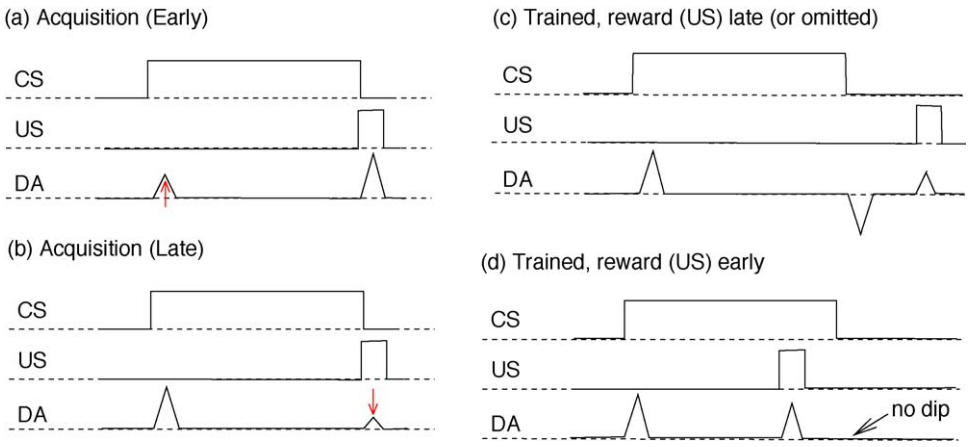

Fig. 1. Schematic of phasic dopamine (DA) recording data. A simple delay conditioning paradigm where a sustained conditioned stimulus (CS) reliably precedes the delivery of a rewarding unconditioned stimulus (US). (a) Early in acquisition, DA initially bursts at the time of US-onset, but then starts spiking at CS-onset as well (up red arrow). There is generally a substantial period of time during training (i.e., hundreds of trials) when bursting is occurring at near maximal rates for both the CS- and US-onset (e.g., Pan et al., 2005). Note that there appears to be little or no evidence for backward-propagating burst firing over training, as predicted by basic versions of the TD algorithm, but not PVLV. (b) Late in acquisition, DA bursting at the time of US-onset progressively starts diminishing so as to eventual disappear entirely (down red arrow). (c) After training, if the US is omitted, or merely delayed as here, a dip in DA firing rates below baseline tonic levels is observed. When the US is then delivered (late), a phasic burst is then seen at that time. (d) If US is instead delivered earlier than expected, a phasic burst in firing is seen. However, no dip is seen at the time the US was originally expected (black arrowhead), implying that (even early) reward delivery has somehow reset the system. *Note*: the 'dip' in panel (c) is shown as roughly symmetrical to the phasic burst activity, but the relative strength of these signals remains somewhat controversial.

central question as to whether the brain actually uses a chaining mechanism like TD, or instead uses error signals that occur *only* at the time of actual stimuli, without recourse to temporal chaining.

Recently, we described a reinforcement learning (RL) model of Pavlovian conditioning based on an algorithm (PVLV, primary value, learned value; O'Reilly et al., 2007) that implies a very specific set of hypotheses regarding the underlying biological substrates driving phasic dopamine cell firing. In this article, we describe some improvements to the original algorithm and provide a more detailed biological interpretation of the overall model than provided in our original paper. Importantly, there is considerable evidence in the empirical literature for the existence of mechanisms very much like those proposed by PVLV, even though it had been inspired by largely computational considerations initially. As a result, we are encouraged that this convergence of biological with computational considerations may mean that PVLV represents a reasonable account of what is actually driving dopamine cells to behave as they do. Reviewing and interpreting much of the relevant empirical evidence is a major purpose of this paper.

An additional goal of this article is to begin to address the second, higher-level question: "How might the mechanisms of phasic dopamine firing—and the widespread dopaminergic signals thereby projected — produce the complex pattern of behaviors characteristic of Pavlovian (and instrumental) conditioning?" Answering this higher-level question is especially difficult for several reasons. For example, dopamine cells project broadly to cortical and sub-cortical areas, and phasic DA firing seems to be fairly indiscriminate with regard to the specificity of the triggering stimuli—approximately 75% of recorded cells consistently fire to unexpected reward regardless of the nature of the reward (Ljungberg et al., 1992; Schultz, 1998). Things are further complicated by the fact that Pavlovian conditioning itself can be decomposed into a complex array of phenomena, with a myriad of conditioned responses, many of which have turned out to be reliably dissociable from one another in lesion studies. Together, these circumstances raise the question, "Exactly what learning is happening where?"

As noted, the computationally oriented literature regarding DA's role in Pavlovian conditioning has heretofore been dominated by the elegant temporal differences (TD) algorithm of Sutton and Barto (Sutton, 1988; Sutton and Barto, 1990, 1998), although there have also been a few non-TD based accounts put forward (e.g., Brown et al., 1999; Izhikevich, 2007; Savastano and Miller, 1998; Tan and Bullock, 2008). Since its development, the basic tenets of TD have been widely adopted by many workers in the field and with considerable theoretical success (e.g., Daw et al., 2003, 2005; Doya, 2002; Houk et al., 1995; Montague et al., 1997; Pan et al., 2005; Schultz et al., 1997; Suri and Schultz, 1999; Schultz and Dickinson, 2000; Suri and Schultz, 2001; Niv et al., 2005). However, for reasons we hope will be made clear, in trying to understand how biological mechanisms might be producing TD-like signals we have been led to think about the underlying processes in terms of more traditional associative mechanisms rather than temporal chaining per se. Nevertheless, as we shall see, at the computational level TD and PVLV share a large common ground, particularly in light of the now widespread acceptance of the TD(λ) framework using eligibility traces which moves TD in the direction of associations instead of chaining.

1.2. Organization of the paper

In the first part of the article, we describe the core PVLV model at a computational level of description, followed by a summary-level biological interpretation. We then proceed to take a more detailed look at the underlying biology by describing six specific hypotheses implied by a biological interpretation of PVLV. This takes the form of a claim-by-claim account which maps several of the key mechanisms of the PVLV model to the underlying biology and examines the empirical evidence that supports each of these individual claims. After that, because of both its importance and complexity, we move on to discuss the role of CS-onset triggered DA bursts in learning generally, focusing on the *conditioned orienting (COR)* and *a* utoshaping responses, both of which are critical to the biological theory. This section also serves to help make sense of a lot of otherwise confusing empirical data.

After that, we describe some new features to the model and provide a specific discussion of how PVLV relates to the widely used TD algorithm, attempting to achieve a loose synthesis of the two approaches. Finally, we conclude the paper with a brief discussion of some predictions that follow from the biological

interpretation of the model and outline areas slated for future development that will serve to broaden the framework to include additional ancillary mechanisms.

2. Overview of the PVLV model

The PVLV algorithm was developed originally to deal with the nuanced demands of several difficult working memory tasks modeled in our broader model of working memory and cognitive control (PBWM; prefrontal cortex, basal ganglia working memory; Frank et al., 2001; O'Reilly and Frank, 2006; Hazy et al., 2006, 2007). In the PBWM model, the basal ganglia act as a dynamic, adaptive gating mechanism for the frontal cortex, to control both motor outputs and working memory updating. PVLV has proved crucial for enabling the basal ganglia in this model to learn this gating function, in the context of the significant temporal credit assignment demands present in tasks that require active maintenance of stimulus information over delays and intervening stimuli. The key motivation was to avoid the strong dependency of the TD algorithm on reliable temporal chains of events in order to span temporal delays: many of the tasks we studied had randomized event sequences, with variable delays between relevant stimuli and reward, that seemed to break the TD chaining mechanism (see Section 6).

In TD, reward prediction errors drive learning about previous states, endowing them with value. When these prior states are subsequently encountered again, their now increased associative value elicits its own prediction error, driving learning about yet prior states. This temporal chaining process repeats so that all reliable states prior to reward are eventually associated with reward, and only the first elicits a prediction error. In contrast, the PVLV algorithm makes use of a simpler *associative* mechanism that accomplishes much the same thing—while using only local information currently available at the synapse at the time of activation.

Specifically, the PV (primary value) component of PVLV learns to associate stimulus cues present at, or predictive of, the time of US (reward) delivery with the expected reward magnitude, and in the process learns to 'cancel' the dopamine burst associated with the US. Mathematically, PV is identical to the Rescorla–Wagner or delta rule algorithm (Rescorla and Wagner, 1972), where it constantly learns to match the actual amount of reward delivered at every time step. We represent the US as the excitatory component of the PV system: PV_e, and the stimulus-driven expectation that cancels this value as the inhibitory component of the PV system: PV_i. The phasic dopamine contribution of the PV system is just PV_e–PV_i.

The LV (learned value) component also learns at the time of US onset (again using a delta-rule formulation), but unlike the PV system, it learns to attach value to any stimulus (CS) that is reliably associated with reward even when this reward occurs later in time. This learning requires some representation of such a stimulus to be active at the time of US-onset in order to establish the necessary contiguity for associative pairing. In this way, subsequent occurrences of the CS come to drive phasic dopamine cell firing *a* t their onset even though the US itself is absent at that time. Critically, the newly acquired LV representations are free to activate when stimulus cues are present in the environment, without the concern that these associations may be unlearned when a CS is present but no US is occurring through most of its duration. This is because the LV system only learns during the time window when an actual US is delivered, or when there is an above-threshold PV expectation of a US that fails to materialize. This is in contrast to the PV system, which is always learning about the instantaneous US value present at every moment, and therefore learns to not expect reward at the time of the CS onset. The LV system also has excitatory and inhibitory components, but we focus more in this paper on the excitatory component since it is the more important piece. The inhibitory component learns to partially cancel the CS-onset phasic bursting, but the timescale is quite slow—typically beyond that of most experiments.

Thus, CSs that precede the onset of the US can drive activation of the LV system, and give rise to the CS-associated dopamine burst. In this way, the PVLV system uses two separate learning systems to account for the two different time periods of dopamine firing: LV at the time of the CS, and PV at the time of the US. In contrast, the TD system accounts for this process through a unitary mathematical formalism of reward prediction and temporal chaining. Although TD is more elegant in this respect, and therefore preferable for certain theoretical treatments, we argue below that PVLV is more consistent with the known biological mechanisms. Furthermore, by virtue of using these two separate mechanisms, PVLV has no dependency on temporal chaining, and is thus robust to variable or unpredictable environments, as contrasted with TD which is typically quite brittle (O'Reilly et al., 2007), but see Daw et al. (2003) for ideas regarding how to make TD more robust to timing. That is, as long as there is some internal representation (e.g., in working memory) of the CS when the US arrives, the LV system will learn about that CS regardless of how long it has been since the CS was presented. For interested readers, we provide a more detailed discussion of these issues in Section 5.

The PVLV model explicitly externalizes the need for active maintenance of information over time, in situations where the relevant sensory information is no longer present in the environment at the time of the reward signal. Thus, PVLV is best considered as one neural element of a larger set of interacting neural systems, exemplified by the PBWM model described above. In PBWM, the prefrontal cortex and basal ganglia are critical for learning to robustly maintain information over time without significant decay, and there is a well-defined set of interactions that enable the complete system to learn about CSs that are no longer present (e.g., in trace conditioning paradigms).

One critical feature of PVLV meriting specific emphasis is that the LV-driven dopamine burst does *not* drive further learning within the LV system itself (recall that learning in LV is contingent on the US being present or the PV expectation thereof); its role is primarily to drive dopamine-modulated learning elsewhere in the brain. This means that as a CS acquires the ability to drive phasic dopamine firing, this same phasic DA signal cannot further self-reinforce the CS–US association that gave rise to it in the first place. Such self-reinforcement would produce a runaway positive feedback cycle that renders the CS driven DA firing impervious to subsequent changes in contingencies (e.g, CS associations would be not extinguishable). That is, once acquired, a CS-onset burst would just keep training itself forever, even when environmental contingencies change. It would also tend to eliminate meaningful differences in value representation between different stimuli. Possible biological mechanisms for this prohibition against self-reinforcement are explored in Section 2.6.

PVLV also serves to synthesize older *associative* explanatory approaches (e.g., Savastano and Miller, 1998) with the more prevalent *reward prediction error* account. PVLV is capable of producing both predictive (i.e., time-specific) as well as non-predictive associations, in the process showing how the former can be produced using only the latter and thus serving to help reconcile the two competing historical accounts. PVLV accomplishes this by moving the treatment of time out of the algorithm itself (which is how TD incorporates time) and instead makes it an explicit neuronal representation—treated like any other neuronal representation. To the extent that temporal cues are reliable associates of reward outcomes, PVLV will learn about them, just like any other reliably associated stimulus. It is noteworthy that Brown et al.

(1999) took a somewhat similar approach, although they incorporated time into their striosomes implicitly, so in this respect that model might be considered more like TD than PVLV.

Finally, we point out that while the PVLV algorithm as described is implemented within a neural network framework (i.e., Emergent), it could also stand alone as a more formal mathematical algorithm much like TD. For readers interested in delving into the formal details of at this time, they are referred to Section 4.3.

2.1. Biological interpretation of the PVLV model

Because of the complexity of the biology involved, we will first provide a summary-level biological interpretation of PVLV in order to help give the reader a basic, general scaffolding upon which to build a deeper understanding. After that, we will provide a more detailed claim-by-claim analysis with the goal of carefully examining the relevant empirical evidence that supports our biological interpretation.

Importantly, the biological interpretation provided is considered to apply across the phylogenetic spectrum from rodents to humans. Accordingly, the evidence examined is drawn from across this spectrum.

A biological-level description of PVLV can be summarized as follows (Fig. 2). The excitatory component of the PV system (PV_e) is associated with the lateral hypothalamic area (LHA) and the inhibitory component (PV_i) with the patch-like cells of the ventral striatum (VS), both of which project to midbrain dopamine nuclei in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc). The LHA represents US value, sending excitatory projections onto dopamine cells, while the VS neurons come to inhibit a dopamine burst that the LHA would otherwise cause once that US becomes "expected." The excitatory component of the LV system (LV_e) is associated with the central nucleus of the amygdala (CNA; Fuster and Uyeda, 1971; Miller, 2000; O'Keefe and Bouma, 1969; Ono et al., 1995; Sanghera et al., 1979; Schultz and Romo, 1990), which can learn stimulus associations to reward outcomes, and drive dopamine firing through excitatory projections of its own to the VTA and SNc.

In a simple Pavlovian conditioning paradigm, these brain areas interact as follows. Prior to conditioning, the occurrence of a reward (US) drives sustained excitatory activity in the LHA (Ono et al., 1986), which in turn drives midbrain DA burst firing via excitatory projections *at US-onset only* (e.g., Floresco et al., 2003; Aston-Jones et al., 2009). These US-driven DA bursts in turn produce two main *intra-PVLV* ("critic") learning effects that change phasic dopamine firing behavior itself. The first involves multi-modal cells of the medial segment of the CNA (mCNA), which are initially responsive only to the US; subsequently, these cells acquire responding also to a CS paired to the US (Ono et al., 1995). As a result, CS-onset also acquires the ability to drive DA bursting via excitatory projections from the mCNA to the midbrain DA system. Note that this crucial link (multi-modal cells of the mCNA) serves to ensure that the population of DA cells driven by the LV_e system (at CS-onset) will be more-or-less the same population driven *a priori* by the PV_e system (at US-onset). *This acquired (learned) ability for CS-onset to drive phasic DA firing is the central mechanism underlying the LV (learned value) system and it can be thought of as purely associative.*

Separately, and in parallel to the learning in the LV_e system, the PV_i system (associated with patch-like neurons of the VS) comes to 'expect' the US by learning about the system's internal state, including temporal representations, immediately prior to US onset. This acquired representation then acts via GABA-ergic projections (and shunting inhibition) to "cancel" the DA spike at the time of reward. Some ventral striatal cells exhibit a ramp up-like activity so as peak at the time of anticipated US-onset (e.g., Apicella et al., 1992). Notably, striosomes have direct, monosynaptic inhibitory projections onto midbrain DA cells (Gerfen, 1985; Gerfen et al., 1987; Smith and Bolam, 1990; Joel and Weiner, 2000), enabling them to cancel (shunt) the excitatory input from LHA, thereby eliminating the DA burst at the time of the reward. This mechanism could potentially also serve to drive dips in DA activity relative to baseline tonic levels when an expected US is omitted or delayed (Fig. 1c). However, there now appears to be compelling evidence that this latter effect may involve projections from the lateral habenular (LH) nucleus to DA cells (Ji and Shepard, 2007; Lecourtier et al., 2008; Matsumoto and Hikosaka, 2007; Shepard et al., 2006). *The acquired (learned) ability for precisely timed ventral striatal representations to inhibit phasic DA firing at US-onset is the central mechanism underlying the PV (primary value) system.*

Finally, as noted earlier the learning that occurs in both the PV and LV systems is restricted to the time of US onset, driven by the global phasic DA signal that occurs at that time (Fig. 3a and b). The critical issue of exactly *how* this learning occurs in PVLV is discussed next.

Each main PVLV subsystem (PV and LV) has its own learning process, largely independent from one another. Combined, they result in the two characteristic changes seen in DA cell firing during Pavlovian conditioning. Importantly, PVLV posits that these two forms of learning occur in separate anatomical locations; therefore, they ought to be readily dissociable by appropriate lesion studies.

- *LV_e learning*: as proposed by PVLV, the crucial LV_e learning occurs in the central nucleus of the amygdala, medial segment, and is the result of the strengthening of synapses from posterior cortical sensory areas, or other amygdalar areas, onto cells of the

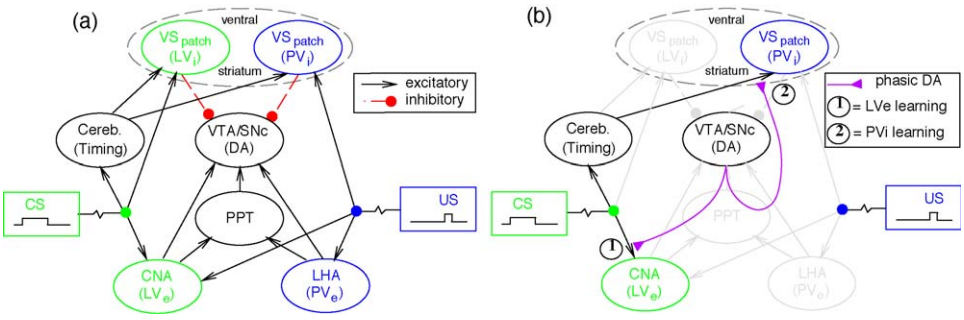

Fig. 2. Biological interpretation of the PVLV algorithm. The PVLV algorithm maps to biological substrates with anatomical and electrophysiological properties capable of playing the roles hypothesized. This figure and discussion is limited to the two main components of the original model: PV system, shown in blue, and the LV system, in green. Updates will be addressed later. (a) Controlling inputs to midbrain DA nuclei (VTA and SNc). Excitatory drive on the midbrain dopamine system comes (primarily) from the lateral hypothalamic area = LHA for USs (PV_e), and central nucleus of the amygdala = CNA for CSs (LV_e). These project directly (and via the PPT = pedunculopontine tegmental nucleus) to the VTA and SNc midbrain dopamine neurons. (b) Two intra-PVLV learning sites emphasized. DA cells shown projecting to two intra-PVLV learning sites (purple): (1) CNA and (2) ventral striatum (VS) so as to drive the learning of LV_e and PV_i, that bootstrap the behavior of the PVLV algorithm itself. See also Fig. 3.

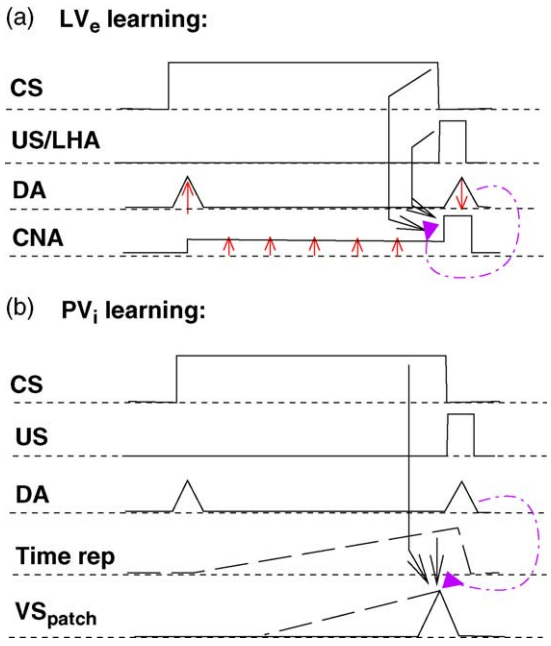

Fig. 3. Two independent learning events change phasic dopamine cell firing during Pavlovian conditioning. Each main PVLV subsystem (PV and LV) has its own critical learning, largely independent from one another, but both learnings occur *at US-onset*. Combined, they result in the two characteristic changes seen in DA cell firing during Pavlovian conditioning. These two learning events occur in separate anatomical locations; therefore, they ought to be dissociable by lesion studies. Phasic DA release promotes learning at both sites (broken purple lines). (a) LV$_e$ learning: according to PVLV, the crucial LV$_e$ learning occurs in the central nucleus of the amygdala (CNA) and is the result of the strengthening of synapses from posterior cortical sensory areas, or other amygdalar areas, onto cells in the CNA that are already responsive to primary reward stimuli (e.g., food and water). (b) PV$_i$ learning: PV$_i$ representations are learned that are *t* imed such that they are able to cancel the excitatory signal to DA cells. PVLV posits the substrate for these PV$_i$ representations to be patch-like MSNs (striosomes) of the ventral striatum. In response to a predictive CS, a subpopulation of ventral striatal cells is known to acquire ramping activation that peaks at the time when the US is expected. By virtue of hypothesized inhibitory projections to midbrain DA cells, these acquired representations can then "shunt" an excitatory signal that comes in when the US occurs, eliminating the DA burst previously seen when the US was unexpected.

mCNA that are already responsive to primary reward stimuli (e.g., food, water; Fig. 3a). Virtually all sensory areas of the cortex are known to project into the CNA and are able to drive sustained firing there (Ono et al., 1995).

As previously emphasized, LV$_e$ learning occurs only at the time of US-onset and requires phasic DA input. The issue of how phasic DA at US-onset can drive learning while its CS-onset counterpart does not is explored later when the specific biological mechanisms are examined in more detail. Note that both pre- (e.g., sensory cortex) and post-synaptic cells (mCNA) are co-active in an overlapping time window, establishing the prerequisite contiguity for *a* ssociative synaptic strengthening (i.e., "Hebb's rule", Hebb, 1949). Note also that this learning is completely *n* on-predictive in the sense that no timing signal is involved whatsoever.

- *PV$_i$ learning*: PV$_i$ representations are learned that are *timed* such that they are able to cancel the excitatory signal to DA cells (Fig. 3b). PVLV posits the substrate for these PV$_i$ representations to be patch-like MSNs (striosomes) of the ventral striatum. In response to a predictive CS, a subpopulation of ventral striatal cells is known to acquire ramping activation that peaks at the time when the US is expected (e.g., Apicella et al., 1992; Deadwyler et al., 2004). By virtue of inhibitory projections onto midbrain DA cells, these acquired representations can then "shunt" the excitatory signal when the US occurs, eliminating the DA burst previously seen when the US was unexpected. The result of this learning is the second (falling) peak illustrated in Fig. 1, line 3. In addition, collaterals projecting multisynaptically via the pallidal nuclei to the lateral habenular nucleus of the epithalamus, by communicating this expectation of a reward, may help that structure produce the "dip" seen when a reward is omitted or delayed (Fig. 1, line 4).

As illustrated in Fig. 3b, we propose a mechanism that depends upon a temporally evolving time representation that provides the critical input to ramping VS MSNs. Synapses from the temporal representations onto VS MSNs are selectively trained by US-onset DA signal. The evolving timing representation is triggered by the onset of the CS—like starting a stopwatch. At the present time, the source of these proposed time representations remain an empirical question. One obvious candidate is the cerebellum, widely thought to be important for representing time (Mauk and Buonomano, 2004). This hypothesis obviously requires that that information be transmitted to the ventral striatum if that is to serve the PV$_i$ role as we propose. There does not appear to be a direct cerebellostriatal connection, but there may be an indirect pathway via cerebellothalamic and thalamostriatal projections (Bentivoglio et al., 1988); alternatively, the evolving time signal may be transmitted from the cerebellum via the frontal cortex (see also Lustig et al. (2005)).

As noted previously, Brown et al. (1999) proposed an alternative to the cerebellum timing mechanism described here, their spectral timing hypothesis that places the stopwatch mechanism directly in the striatum, positing cells specialized for timing. In any event, the exact nature of the timing mechanism remains an empirical issue, and either could provide a sufficient representational substrate to enable PV$_i$ learning.

Again we emphasize that, for *both* sites, learning occurs only around the time of *US-onset*. Also, note that the ramping pattern of VS cell activity (dashed line) can be explained as an effect of stimulus generalization with regard to an evolving time representation, which will be progressively less similar to its nature at the time of US-onset as one moves backward in time (Mauk and Buonomano, 2004). Finally, the restriction of burst firing to only the *onset* of a US or CS is proposed to be an intrinsic property of the excitatory projection pathway into the DA nuclei. PVLV does not generally require that any learning occur at DA cell synapses themselves; the strength of these synapses is thought to be relatively fixed, either hardwired during ontogeny, acquired during early development, or some combination thereof. A recent paper provides partial support for this idea of relatively little plasticity in that, while the authors reported a temporary facilitation of glutamatergic synapses onto DA cells, the effect did not appear to convert to L-LTP (Stuber et al., 2008).

As a final note for this section remember that, in addition to the intra-PVLV learning just described, phasic DA firing also affects many *extra-PVLV* sites (i.e., outside of PVLV proper), especially the striatum and cortex, that collectively produce the broad set of behavioral phenomena recognized as Pavlovian and instrumental conditioning. In our models of this broader system, phasic DA trains striatal "Go" and "NoGo" representations to learn to facilitate the most adaptive actions and to suppress less adaptive ones in particular stimulus contexts (Frank, 2005; O'Reilly and Frank, 2006). The separate categories of intra- and extra-PVLV learning sites can be loosely mapped to the "critic" and "actor" roles common in the reinforcement learning framework more generally.

2.2. Claim-by-claim review of PVLV-proposed mechanisms of phasic dopamine firing

Hopefully, the reader now has a basic general understanding of PVLV as a biological model. We now proceed to a detailed

Table 1
Six core biological hypotheses implied by the PVLV algorithm.

#1: PV_e representation $\approx$ LHA
#2: PV_i representation $\approx$ patch-like MSNs in the VS
#3: LV_e representation $\approx$ CNA
#4: CNA multi-modal cells are not trained by CS-triggered DA
#5: With overtraining, LV_i partially mitigates phasic DA firing to CS-onset
#6: Phasic DA bursting arises from sustained inputs

claim-by-claim examination of the empirical evidence supporting this interpretation. Clearly, the validity of PVLV as a *biological* model hinges on the identification of functional neuronal substrates corresponding to its core components: the PV and LV systems. In this section, we examine the biological evidence for these separable systems in more detail, outlining the empirical data bearing on each of six key hypotheses implied by the PVLV framework. (See Table 1 for summary.)

2.3. PVLV hypothesis #1: PV_e representation $\approx$ LHA

- *Synopsis:* The lateral hypothalamic area is the main site for the *reactive* representation of US value for rewarding stimuli such as food, water, etc., and this provides the main excitatory signal driving phasic dopamine bursting after primary reward onset.

This is a relatively widely accepted hypothesis with direct, well established data. Cells of the LHA receive direct projections from primitive sensory areas associated with primary reward (e.g., gustatory, Fukuda and Ono, 1993; Norgren, 1976), and respond to the occurrence of reward with *sustained* firing (e.g., Schultz and Romo, 1990; Mora et al., 1976; Nakamura and Ono, 1986; Ono et al., 1981; Rolls et al., 1976), with little evidence for attenuation after stimulus onset (Nakamura and Ono, 1986; Ono et al., 1986). Nor is there evidence for attenuation after conditioning (Nakamura and Ono, 1986; Ono et al., 1986). The LHA sends excitatory glutamatergic projections directly to the midbrain dopaminergic nuclei (VTA & SNc; Phillipson, 1978), and even more densely to the pedunculopontine tegmental (PPT) nucleus (Semba and Fibiger, 1992), which in turn sends both glutamatergic and cholinergic projections to the midbrain DA nuclei (Floresco et al., 2003; Semba and Fibiger, 1992).

2.4. PVLV hypothesis #2: PV_i representations $\approx$ patch-like neurons in the VS

- *Synopsis:* Patch-like GABAergic neurons in the ventral striatum are the main substrate for the learned representation of a *US expectation* (PV_i). Inhibitory projections therefrom to midbrain DA cells shunt excitatory inputs thereby eliminating the phasic burst for the US. The VS also projects via the pallidum to the lateral habenular (LH) nucleus of the epithalamus, helping that substrate to compute when an expected reward has been omitted.

During conditioning, midbrain dopamine neurons eventually stop firing a phasic burst at the time of a reward (Fig. 1b), and if an expected reward is then omitted (or delayed), there is a phasic pause or dip in tonic firing at the expected time (Fig. 1c). These effects are thought to be global across the majority of DA neurons in the VTA and SNc (e.g., Schultz, 1998). Under the PVLV framework, these functions are subserved by the PV_i subsystem. We argue that patch-like neurons (*striosomes*) in the ventral striatum appear to be particularly well suited for such a role. For reasons that will become clear, we divide the discussion into two parts: the blocking of phasic US bursting, and the production of a dip in the otherwise tonic DA firing associated with omitted expected rewards.

2.4.1. Blocking phasic US bursting

The evidence in favor of ventral striosomes (patch-like cells) mediating the blocking of US dopamine bursts rests upon four interrelated lines of empirical evidence. Together, these provide what we think is a compelling, but still somewhat circumstantial case—hopefully more direct tests will be conducted. First, a subpopulation of medium spiny neurons (MSNs) of the ventral striatum exhibits acquired ramping activity after Pavlovian conditioning that peaks at the time of US-onset, providing the proper timing for blocking US bursts. Second, striosomes in the *dorsal* striatum have monosynaptic GABA-ergic inhibitory projections directly onto DA cells. Third, MSNs of the *ventral* striatum are also histologically divisible into two cell populations—patch-like vs. matrix-like, though they are not as anatomically segregated as in the dorsal striatum. Fourth, ventral (but not dorsal) striatum projects to both the dorsal and ventral tiers of DA cells, i.e., to potentially all DA cells of both the VTA and SNc. We address each of these points in turn.

Multiple groups have described cells in the ventral striatum that, after conditioning, acquire a ramp-like pattern of activity that is triggered by a conditioned stimulus (with considerable variability in lag period), and which peak immediately preceding the expected occurrence of a reward (Schultz et al., 1993b; Cromwell and Schultz, 2003; Deadwyler et al., 2004; Roitman et al., 2005). For example, Schultz et al. (1993b) found that approximately 10–15% of the total population of recorded cells acquired an anticipatory pattern of activity that began some time (with variable lag) after the onset of one stimulus (e.g., instruction, trigger) and progressively ramped up so as to peak at the anticipated time of a second stimulus (e.g., trigger, reward; Schultz et al., 1993b; Cromwell and Schultz, 2003). Of this population, the great preponderance of cells were anticipatory of reward (e.g., 43/60, Schultz et al., 1993b). In addition, some ventral striatal neurons fire immediately prior to primary rewards with no ramping (Schultz et al., 1993b; Cromwell and Schultz, 2003; Deadwyler et al., 2004), critically including at the expected time of reward when no rewards are actually presented (i.e., in extinction trials). Identification of both these subpopulations as MSNs and not TANs (tonically active neurons) was made on the basis of electrophysiological signatures. However, it has not heretofore been possible to specifically characterize these cells as striosomes on the basis of electrophysiology; we make that link on the basis of anatomy as follows.

The caudate/putamen (*dorsal* striatum) exhibits two distinct histological/histochemical compartments throughout its extent, divisible into island-like *patches* (containing striosomes; approximately 20% by volume) and an intervening *matrix* (containing matrisomes; 80%) everywhere else (Gerfen, 1985; Graybiel, 1998; Joel and Weiner, 2000). Matrisomes have been extensively implicated in both the direct and indirect pathways thought to gate motor activity in the frontal cortex (Brown et al., 2004; Frank, 2005; Frank et al., 2001, 2004; Mink, 1996; Surmeier et al., 2007), while the identification of a functional role for the patch compartment's striosomes has remained elusive (e.g., Joel and Weiner, 2000). One critical difference between matrisomes and striosomes is that the striosomes project via $GABA_B$-ergic synapses directly onto cell bodies and proximal dendrites of ventral tier DA cells of the SNc (Joel and Weiner, 2000). In contrast, matrisomes project via $GABA_A$-ergic synapses predominantly onto $GABA_A$-ergic interneurons in the substantia nigra, pars reticulata (SNr).

The ventral striatum also has two distinct sub-populations of medium spiny neurons, which have been described as patch-like and matrix-like because of the histological staining characteristics

they share with their dorsal counterparts (Joel and Weiner, 2000). The patch-like cells are especially prevalent in the shell of the nucleus accumbens in rats (Gerfen, 1985; Gerfen et al., 1987). However, the VS (especially NAc) does not exhibit the histological compartmentalization of these cell types seen in the dorsal striatum, which has made it difficult to establish connectivity differences between these cell types similar to those established for the dorsal striatum. Nonetheless, two different subpopulations of MSNs have been described in the NAc based on connectivity, one projecting onto DA cells of both the VTA and SNc and the other synapsing onto GABAergic neurons of the SNr (Berendse et al., 1992; Groenewegen et al., 1990; Joel and Weiner, 2000). Thus, while we do not yet know definitively that it is the patch-like cells in the ventral striatum that form $GABA_B$-ergic synapses onto DA neurons, we do know that such connections from VS to DA cells exist, and it is likely that they would originate from these neurons, given their other commonalities with dorsal striosomes. Indeed, based on these same considerations, Joel and Weiner (2000) were led to hypothesize a similar functionality to that proposed here.

The final data point concerns the ability of projections from ventral, but not dorsal, striatum to modulate the entire population of DA neurons in the VTA and SNc, making it more plausible that the PV_i-like mechanism is associated with ventral striatum. Specifically, DA cells fall into two distinct sub-populations based on several histological, electrophysiological, and biochemical characteristics: *dorsal tier* and *ventral tier*. The former predominate in the VTA, but also occur with lower density in the SNc; the latter have a reciprocal distribution: highest in the SNc, but also definitely existing at lower density in the VTA (Joel and Weiner, 2000). The MSNs of the VS project to *both* tiers (i.e., the entire VTA and SNc) putting them in a position to inhibit burst firing globally (Joel and Weiner, 2000). In contrast, MSNs of the dorsal striatum project only to ventral tier dopamine cells (i.e., mostly SNc), and with a more focused topographical pattern (Joel and Weiner, 2000).

In summary, all of the necessary pieces of evidence are in place to suggest that ventral striatal patch-like neurons can block phasic US DA bursts, but these pieces have not yet been put together in a single study to definitively test our PV_i hypothesis. We propose various testable predictions for such studies in the predictions section of the general discussion.

We can contrast this proposal with several others in the literature for how US-driven dopamine bursts are cancelled. For example, some have proposed that it may be dorsal striosomes (instead of ventral) that serve to inhibit phasic dopamine firing (Houk et al., 1995). However, as noted above, dorsal striatum projects rather locally and topographically to the SNc, and only ventral tier DA cells (Joel and Weiner, 2000), making it less suitable for a global US cancelling mechanism. Alternatively, Wickens and Kotter (1995) proposed the ventral *matrisomes*, but these synapse preferentially onto midbrain GABA-ergic interneurons and therefore serve to transmit a quasi-excitatory signal (via disynaptic indirect disinhibition), not an inhibitory one (Joel and Weiner, 2000; Gerfen, 1985). Consistent with this anatomy, electrophysiological data shows that matrisomal stimulation can produce excitation of DA cells (via inhibitory release), not inhibition (Grace and Bunney, 1979, 1985, 2000). Finally, Brown et al. (1999) proposed a combination of dorsal and ventral striosomes for the US-cancelling role, which is closer to our own thinking in this respect, though important differences exist as to the nature of the timing signal as mentioned earlier, and they also propose that the striatum is also the source of the acquired excitatory signal (LV_e), not the CNA as we propose.

2.4.2. Generating dips for omitted expected rewards

We originally made the parsimonious assumption that the blocking of phasic burst firing at US-onset and the dip in tonic firing seen when an expected US is omitted were both manifestations of a single inhibitory mechanism. However, recent findings that the lateral habenular nucleus of the epithalamus can produce pauses in DA cell firing (e.g., Heldt and Ressler, 2007; Ji and Shepard, 2007; Lecourtier et al., 2008; Matsumoto and Hikosaka, 2007; Shepard et al., 2006) suggest that it may be important for generating the dip associated with omitted rewards. Nevertheless, we continue to believe that the VS plays the dominant computational role in predicting the occurrence of reward, and serves as a driving input to the lateral habenular nucleus.

Specifically, our model posits that VS neurons send a net excitatory (disinhibitory) signal to the habenular neurons, whereas primary rewards (USs) inhibit the habenula (Matsumoto and Hikosaka, 2009), such that an expected but omitted US would provide net excitation of the habenula, causing in turn an inhibition in the tonic firing of DA neurons. Aversive stimuli may directly activate the habenula to produce dips in dopamine firing that have been observed under such conditions (Matsumoto and Hikosaka, 2007).

Other considerations also suggest that there may be a different neural substrate for blocking US bursts and generating dips in tonic firing. For example, to be effective in cancelling a US-driven DA burst, shunting inhibition would need to occur slightly *before* any excitatory signal that would otherwise produce the burst. In contrast, the dips associated with omitted expected rewards occur clearly *after* the expected US timing, and the duration of these dips extends significantly longer than corresponding bursts (e.g., Hollerman and Schultz, 1998; Bayer et al., 2007; Satoh et al., 2003; Roesch et al., 2007). In addition, there is now a growing body of evidence suggesting that tonic dopamine firing and phasic bursts superimposed on tonic levels are driven by two dissociable networks (Floresco et al., 2003). For example, tonic firing can be controlled by the ventral pallidum (VP), while phasic bursting can be induced by stimulating the pedunculopontine tegmental nucleus (PPT), but only in DA cells that are already tonically firing (Floresco et al., 2003). Finally, the inhibitory demands of shunting a US burst and generating an actual dip or pause in tonic firing may be quite different, assuming that the tonic firing operates through intrinsic mechanisms in the DA neurons. Interestingly, one does not see a dip in tonic firing just prior to the expected time of US onset, despite the fact that the VS neurons are exhibiting a ramping-up of activation—this is consistent with a shunting-like mechanism that does not affect tonic firing rates.

The evidence in support of the habenula playing a specific role in pausing tonic DA firing is becoming increasingly compelling. For example, direct electrophysiological stimulation of this nucleus produces pauses in tonic DA firing (Shepard et al., 2006), and this effect appears to be mediated via glutamatergic projections onto $GABA_A$-ergic interneurons of the VTA and SNc (Shepard et al., 2006). Midbrain inhibitory interneurons are known from other studies to play a role in controlling tonic DA firing (Floresco et al., 2003). Very recent data show that primary rewards produce an *inhibition* of habenular cells (Matsumoto and Hikosaka, 2009). This inhibition would then prevent the habenular cells from becoming active and inhibiting DA tonic (and thus phasic) firing. Importantly, this suggests that the habenula is likely *not* involved in the shunting of phasic DA burst firing, because that would presumably require it to become active for primary rewards, not inhibited. This is consistent with our proposal for a division of labor and two separate systems for the shunting of phasic bursting versus the production of phasic pauses in tonic firing.

Among the sources of major afferent projections to the lateral habenula are the lateral hypothalmic area and the output nuclei of the basal ganglia, including the internal segment of the globus pallidus (GP_i) and the ventral pallidum (Shepard et al., 2006; Herkenham and Nauta, 1977). This later pathway is consistent

with the idea that the ventral striatum may be critical in contributing to lateral habenular signaling, as suggested above. It is possible that the same patch-like VS neurons proposed to cancel US bursts via shunting inhibition of DA cells also contribute to the phasic DA dips for omitted expected rewards via collaterals producing habenular disinhibition. Alternatively, perhaps the non-ramping subpopulation are more important for the omitted rewards component, due to the absence of ramping-like effects on tonic DA firing levels prior to expected rewards. In either case, the information that an expected reward has been omitted could be transmitted by the *a* bsence of a usual reward signal from the LHA, an area known to project to the habenula as noted above.

To summarize our complete story about the PV_i mechanism: the patch-like neurons in the VS provide a shunting inhibition directly onto midbrain DA cells, preventing US-onset burst firing, while leaving tonic activity intact. In addition, these same neurons and/or another population of VS neurons send a net excitatory signal to the habenula, which is inhibited by primary rewards from the LHA, but produces net activation when such rewards are omitted. This activation then transiently inhibits DA tonic firing to produce the pauses seen.

2.5. PVLV hypothesis #3: LV_e representation $\approx$ CNA

- *Synopsis:* Multi-modal glutamatergic projection neurons of the medial segment of the CNA are the substrate for the *a* cquired representation of CS reward value for the purpose of driving phasic dopamine bursting at CS-onset.

Recall that the proposed role for the LV_e subsystem is to learn to represent value for CSs predictive of reward, and then to drive phasic DA bursts CS-onset. Glutamatergic projection neurons of the mCNA appear to be well suited for this role. We are not the first to suggest the CNA, or the amygdala generally, for such a role (e.g., Fuster and Uyeda, 1971; O'Keefe and Bouma, 1969; Miller, 2000; Sanghera et al., 1979; Schultz and Romo, 1990).

The case for the identification of LV_e with the CNA rests upon four interrelated lines of empirical evidence. First, the amygdala is crucial for positively-valenced reward conditioning, in addition to its better-known role in fear conditioning. Second, the CNA, but not the basolateral amygdalar complex (BLA), projects to midbrain DA nuclei, and CNA neurons have been shown to drive DA bursting. Third, multi-modal cells of the CNA (medial segment), initially responsive to reward, acquire responsiveness to CSs that predict that same reward. Fourth, lesions of the CNA selectively impair the acquisition of CS-generated CRs (autoshaping and conditioned orienting responses), but not US-generated CRs, consistent with a loss of CS-driven phasic DA bursting. We review each in turn.

While the amygdala has long been associated with fear conditioning (see, e.g., LeDoux, 2003, for review), it is now well established that both the BLA and CNA also code for positively valenced stimuli (Baxter and Murray, 2002; Belova et al., 2007; Ono et al., 1995; Paton et al., 2006; Schoenbaum et al., 1999, 2003). Stimuli that might be thought of as unconditioned stimuli, i.e., stimuli with reliable ecological value, seem to have a kind of innate hardwiring to specific amygdalar cells for either positive or negative valence, with the relative proportion favoring negatively over positively valenced cells approximately two-to-one (Schoenbaum et al., 1999).

At a rough functional level, the amygdala can be divided into a cortex-like *basolateral* set of nuclei (BLA; basal, lateral, accessory basal nuclei), and a more striatum-like *central* set of nuclei (CNA; medial segment, lateral segment) (Amaral et al., 1992; Pitkanen, 2000). Both BLA and CNA receive broad projections from all over the cortex, with the CNA receiving such projections both directly, and via a kind of funneling pattern from the BLA. Afferents to the BLA appear to be considerably more extensive and of a finer granularity, however (Amaral et al., 1992). The medial segment of the CNA sends glutamatergic (excitatory) projections to the PPT, which in turn projects to midbrain DA cells (Fudge and Haber, 2000; Wallace et al., 1992; Takayama and Miura, 1991), but the exact nature of these projections remains somewhat unclear. mCNA also sends some excitatory projections directly to the midbrain DA nuclei as well (Fudge and Haber, 2000; Wallace et al., 1992), and also to the LHA (Petrovich et al., 2001, 2002). In contrast, the BLA does *not* project independently to the DA midbrain areas, only indirectly doing so via its projections to the mCNA. Finally, several electrophysiological studies have provided strong corroboratory evidence in that stimulation of CNA neurons can cause dopamine cell firing in the VTA and SNc and/or DA release in target areas (Ahn and Phillips, 2003; Rouillard and Freeman, 1995; Fudge and Haber, 2000).

Multi-modal neurons of the CNA respond in a sustained manner to primary reward (USs). In Pavlovian paradigms these same cells (initially responsive only to a US) then learn to fire also for an associated CS (Ono et al., 1995). Furthermore, consistent with the idea that L-LTP (late, or permanent, long term potentiation) is occurring for these events, immediate early gene expression has been observed in CNA cells, particularly those that project to SNc, in response to a visual stimulus predictive of reward (Lee et al., 2005).

The final evidence for the CNA playing the critical role in driving phasic DA bursting at CS-onset comes from studies showing that CNA lesions interfere with a set of Pavlovian conditioned responses (CRs) that are likely to depend on CS-driven phasic DA. Specifically, the *conditioned orienting response* (COR) and *autoshaping* are selectively affected by CNA lesions, leaving many other CRs intact. Critically, these two CRs are natively elicited by the CSs prior to conditioning (as contrasted with Pavlov's iconic canine salivation, which is elicited innately by a food US but by the CS only after conditioning). For example, a bright localizable light stimulus elicits an *unconditioned* orienting response prior to any CS–US pairing. Thus, these behavioral activations are precisely timed so as to be uniquely trainable (i.e., the orienting response can be potentiated) by a phasic DA burst triggered at CS-onset (once conditioning has occurred). We discuss this important issue in greater detail in the next major section of the paper following this claim-by-claim analysis.

2.6. PVLV hypothesis #4: CNA multi-modal cells are not trained by CS-triggered DA

- *Synopsis:* The critical learning in CNA occurs at synapses between incoming CS sensory representations and multi-modal cells initially responsive to US, and these are trained by phasic DA at the time of US-onset *only*. CS triggered phasic DA signals do *not* train associations in the CNA.

Consistent with the PVLV model, there is strong evidence for relevant plasticity in the CNA, specifically for multi-modal cells in the medial segment to acquire the ability to also fire in response to CSs paired with a US (Ono et al., 1995), and for early gene expression characteristic of L-LTP (Lee et al., 2005). For this plasticity to function properly, the PVLV model predicts that DA bursts generated by the CS onset itself should not be capable of driving plasticity within the CNA (see earlier computational section for a theoretical explanation).

It is well established that DA cells of the SNc and VTA project to the amygdala, most densely to the mCNA (e.g., Fallon and Ciofi, 1992; Amaral et al., 1992). Furthermore, DA blockade in the amygdala has been shown to impair LTP and learning (Andrzejewski et al., 2005). Finally, and most importantly, the acquisition

of phasic bursting by DA cells to a compounded CS has been shown to be prevented in traditional blocking paradigms (Kamin, 1969a,b) that prevent US-onset DA firing by pretraining on the blocking CS (Waelti et al., 2001). Together, these findings provide compelling evidence that phasic DA is important for plasticity in CNA. However, why should phasic DA at the time of US-onset drive learning, but CS-onset DA signals not, as we suggest based on computational considerations? We see two possible mechanisms for this at the biological level.

A mechanism that we currently consider most likely is that the time constants involved are such that any self-generated phasic DA signals would come too late to be effective in self-training (see Tan and Bullock, 2008 for an earlier, similar proposal, albeit in the striatum). That is, glutaminergic activation of post-synaptic receptors must be ongoing for a some threshold period of time before subsequent DA receptor activation can exert its modulatory effects on learning. Only for the case of US-onset DA release after an initial CS onset would the glutamate receptors have been active long enough to drive plasticity. Interestingly, this account predicts that CS conditioning would require that the CS signal precede the US by some minimal amount of time (e.g., 200 ms). Indeed, a temporal ordering constraint has indeed been long recognized in the behavioral literature (Schmajuk, 1997; Schneiderman, 1966; Smith, 1968), and electrophysiological data supports a temporal ordering constraint with regard to the effects of dopamine on learning (Bao et al., 2003).

A second possibility is that some intra-amygdalar network mechanism might be responsible for blocking CS-driven DA plasticity effects. For example, the lateral CNA (latCNA) contains GABA spiny neurons much like the striatum (Davis et al., 1994), and it projects to the mCNA. Thus, it could inhibit any plasticity effects from self-generated dopamine bursts.

Finally, a third possibility deserves mention. It is that the CS–US pairing in the mCNA is not trained by phasic dopamine at all. Instead, the occurrence of the US itself may act alone as the teaching signal to produce the association. This, in fact, is the mechanism closest to how the PVLV algorithm actually implements this constraint since it is the simplest. Consistent with this possibility, CNA neurons responds natively to USs (e.g., Ono et al., 1995), and the CNA and the LHA (where US representations are known to be encoded) are reciprocally interconnected (Nakamura and Ono, 1986; Ono et al., 1986). Thus, LHA inputs alone may serve to train the CNA neurons. Going against this possibility, of course, is the dense dopaminergic innervations of the amygdala and the blocking result cited above (Waelti et al., 2001).

Overall, then, we think that phasic DA is important in the CNA, subject to the temporal constraint that CS-generated DA signals do not act quickly enough to produce self-training.

2.7. PVLV hypothesis #5: with overtraining, LV_i partially mitigates phasic DA firing to CS-onset

- *Synopsis:* With overtraining, phasic DA firing to CS-onset is significantly reduced, but persists indefinitely. However, if there is a CS_2 prior to, and predictive of, CS_1, DA firing to CS_1 is eliminated. These effects can be explained in terms of a CS-activated inhibitory representation (LV_i) that inhibits the excitatory influence from the LV_e (CNA).

One of the underrecognized characteristics of the acquired phasic DA firing to CS-onset is its behavior in the face of overtraining. On the one hand, it appears to show remarkable persistence over tens of thousands of trials (Ljungberg et al., 1992). On the other, however, the rate of responding does go down significantly with training, ending up with a burst firing rates on the order of 50–75% of its earlier peak rate (Ljungberg et al., 1992). Thus, there appears to a kind of partial habituation of the effect that occurs with overtraining, even absent any obvious predictive cue of CS occurrence. Note that it makes sense ecologically that some habituation should occur, particularly in a recurring context that might serve to lower its motivational value. On the other hand, it also makes sense that it should not go away completely, since it could very well be a life-saving signal at another place and time, or under other circumstances. Thus, we interpret this partial habituation-yet-persistence pattern as a compromise between these competing ecological influences.

We have found from a computational perspective that having a very slow learning process that acts to partially cancel CS-onset DA firing (just like PV_i cancels US onset DA firing) is beneficial for learning complex working memory tasks (O'Reilly and Frank, 2006). This enables the system to be sensitive to changes in CS-reward associations, rather than just raw values, by providing an adaptive baseline against which the current CS-reward value is compared. In addition, this partially canceling is critical for unlearning in extra-PVLV ("actor") sites like the striatum because if a CS always triggers a maximal DA burst then one gets too much Go learning instead of the balance needed for flexible behavior.

Implementationally, the PVLV algorithm includes an LV_i mechanism that is identical to the learning in the LV_e mechanism, just at a slower learning rate, and which provides an inhibitory signal to the dopamine system. Biologically, a slow-learning mechanism such as this could perhaps be located in the lateral CNA (latCNA), with its GABA-ergic projection neurons (Davis et al., 1994). Alternatively, an LV_i-like mechanism could exist in the ventral striatum and work much like the PV_i, except in its being triggered by a CS_2 (either explicit, or implicit/contextual). Indeed, Schultz et al. (1993b) found cells in the VS that fit this exact pattern: subpopulations of VS neurons exhibit peaking at the time CS-onset (Schultz et al., 1993b; Cromwell and Schultz, 2003), triggered by the occurrence of a still prior stimulus. In that case, the prior CS_2 was explicit. It is easy to imagine, however, that animals could develop LV_i representations from implicit/contextual signals as well – such predictions would be less exact and, therefore, the mitigation of dopamine firing only partial, which is what is seen empirically with overtraining and no explicit CS_2 (Ljungberg et al., 1992).[1]

2.8. PVLV hypothesis #6: phasic DA bursting arises from sustained inputs

- *Synopsis:* DA firing to CS and US inputs is invariably in the form of a phasic burst, even when these input signals persist for sustained periods of time. The pedunculopontine tegmental nucleus (PPT) or the midbrain DA nuclei themselves appear likely to be responsible for producing this bursting property, in a manner consistent with the temporal derivative $\dot{Y}$ mechanism now used in PVLV.

Both purported excitatory drivers of phasic DA firing, the lateral hypothalmic area and the central nucleus of the amygdala, fire in a *sustained* manner to appropriate stimuli. That is, so long as the

[1] It is interesting that autoshaping CRs, which are typically immune to being blocked (Tomie, 1981; see discussion later) can, in fact, sometimes be blocked, as was also explored by Tomie (1981). He and others had found that prior exposure to a CS, not contingent on reward, caused the retardation of subsequent acquisition of autoshaping CRs, a phenomenon usually termed *l* atent inhibition. Tomie (1981) was able to show that, at least for the autoshaping CRs he studied, this was due to what he called the *context-blocking effect*. By moving the subjects to a new context he eliminated the retarded acquisition, that is, eliminating any blocking effect. Presumably, the familiar context was acting as a predictive CS_2 that could serve to shunt the CS_1-onset DA burst. This is the basic idea underlying the second mechanism for LV_i just discussed, and for the experiment we propose later to explore this issue.

underlying stimuli are present in the environment, they remain active. On the other hand, DA cells appear to burst-fire only at stimulus *onset*. This raises a critical question: Why do DA neurons not continue firing at their very high phasic rate throughout the duration of the stimulus? In other words, what turns them off so crisply after their initial burst?

Based on our review of the literature, it appears that the answer to this question remains unresolved at the present time. However, several promising candidate mechanisms are suggested by available evidence. For example, this pattern of firing might be caused by internal dynamics within the pedunculopontine tegmental nucleus (PPT) as argued for example by Brown et al. (1999), or within the VTA/SNc nuclei themselves. For example, excitatory projections to the midbrain DA cells might have collaterals that synapse onto nearby GABA-ergic inhibitory interneurons producing a sustained inhibitory signal that lasts as long as the input.

Alternatively, a combination of direct LHA-/CNA-to-DA excitatory projections followed immediatley by PPT inhibitory (e.g., cholinergic) projections could be responsible. Or, intracellular dynamics within DA neurons themselves could play a role. Some support for the latter mechanism is provided by the finding that metabotropic glutamatergic receptors (mGluR) on DA cells in the midbrain nuclei generate IPSPs that follow the immediate EPSPs produced by the ionotropic Glu receptors (Fiorillo and Williams, 2000). In any event, whatever the mechanisms, it has to continue to block the re-emergence of phasic bursting throughout the duration of a sustained excitatory input. This would seem to favor some sort of sustained inhibitory collateral signal as perhaps the most likely candidate. Any of these mechanisms is consistent with the $\dot{Y}$ (“Y-dot”) temporal derivative (Sutton and Barto, 1990) used in the current version of PVLV.

3. The role of CS-onset phasic DA signals in conditioning and learning

Having addressed the specific mapping of PVLV onto the underlying biology in the previous section, we now turn to the larger question of how phasic dopamine signals shape the learning of macroscopic behaviors. As we have described, all of the *intra-* PVLV (“critic”) learning is driven by US-onset DA firing; with regard to *extra-* PVLV (“actor”) learning, however, the burden is shared. Indeed, we see a nice complementarity between the roles of US- versus CS-onset DA in shaping macroscopic behavior. In a nutshell, US-driven DA plays predominantly an *after-the-fact* reinforcement role akin to Thorndike’s Law of Effect (Thorndike, 1911). On the other hand, CS-onset DA plays a *preparatory* role that serves to train up representations of goal states and exploratory behaviors associated with the specific *opportunity* signaled by the CS. These opportunity-specific representations can then serve to activate the subject generally and promote a trial-and-error exploratory search among various behaviors, with subsequent US-onset DA signals then selecting the most effective and efficient actions. Under this framing, Pavlovian mechanisms can be seen to be important enablers of *instrumental* learning, with the onset of a particular CS signaling the beginning of a window of instrumental opportunity.[2] In our original paper (O’Reilly et al., 2007), we described how PVLV could simulate several of the phenomena most associated with Pavlovian conditioning: blocking; over-shadowing and summation; conditioned inhibition, and second-order conditioning. With the exception of the last one, all of these classic phenomena are due to effects associated with US-onset DA firing. Because of its equally critical role, we focus here on the effects of the phasic DA signal driven by CS onset, which is supported by the LV system in PVLV, and the CNA in our biological account (Figs. 1 and 2).

Importantly, we might expect that the behavioral footprints left behind by CS-driven DA ought to be more distinctive than those associated with the US, because the latter is somewhat over-determined: many different forms and sites of learning are driven by US signals, and these are reliably accompanied by DA and other neuromodulatory signals quasi-innately. In contrast, CS-driven DA has its own learning curve, which should be reflected in whatever further learning depends upon it. Furthermore, by definition the CS starts out as a “neutral” stimulus in the current context with many fewer pre-learned associations and/or affordances. For these reasons, the behavioral consequences associated with CS-triggered DA learning ought to quite distinctive.

We focus primarily on the subset of conditioned responses (CRs) that we claim are best explained as being dependent on CS-onset DA firing. As reviewed by Holland (1984) in a particularly informative paper, these CRs are behaviors that are natively generated *by the CSs themselves*, albeit sometimes in weak or even latent form (e.g., Wasserman, 1981). This subset of CS-native CRs is readily dissociable from other CRs that are natively generated by the USs (Dykman, 1965; Gallagher et al., 1990; Holland, 1984). Critically, the acquisition of the CRs in question are also known to be dependent upon an intact striatum, the most important locus of phasic DA signaling according to our thinking and many others. This striatal dependency is critical for tying the *timing* of phasic DA signaling to the anatomical *location* of its action.

The two instances of CS-onset DA dependent CRs that we will focus on have been extremely well studied and are known as the *conditioned orienting response* (COR), and *autoshaping* (also termed “conditioned approach to the CS” or “sign tracking”). After first describing each briefly, we will then explain why both of these phenomena should be driven by CS-onset phasic DA according to the PVLV model, which is consistent with the well documented effects of CNA lesions in these cases. Then, we go on to account for two other related phenomena in this domain that are consistent with the PVLV model. These include the absence of blocking effects in autoshaping, and the acquisition of the generalized form of Pavlovian instrumental transfer (PIT).

3.1. The conditioned orienting response (COR)

Animals typically respond to a novel CS with an *orienting response (OR)* that is highly stereotyped per species and also specific to the CS involved (e.g., Pavlov, 1927; Wasserman, 1981; Holland, 1984). For example, in response to a novel visual stimulus (e.g., light), a rat will rear on its hind legs for a short period of time, before returning to whatever it had been doing prior to the initiation of the light. In contrast, a novel (non-aversive) auditory CS will elicit a startle OR, which is a jump that is morphologically distinguishable from the rearing described for visual CSs.

With repeated exposure to a particular CS, the OR it elicits habituates (Pavlov, 1927; Sokolov, 1963; Gallagher et al., 1990). However, if a CS is paired with a US, animals reacquire the OR during training in a characteristic U-shaped pattern that reflects this habituation-reacquisition sequence (Hatfield et al., 1996; Gallagher et al., 1990). This rescue of the orienting behavior from habituation is thus called the *c* onditioned orienting response, or COR (Holland, 1977, 1984; Gallagher et al., 1990; Gallagher and Holland, 1993, 1994; Hatfield et al., 1996; Han et al., 1997; Holland et al., 2002; Groshek et al., 2005).

[2] This is not to say that CS-onset driven DA signals cannot train behaviors immediately prior to their occurrence—this is exactly the mechanism that trains the conditioned orienting and autoshaping CRs as we describe below. The point is that the ecological value of such learning is served in proportion to its contribution to ultimately achieving some *primary* reward as defined by the instrumental opportunity.

The COR is dependent on an intact central nucleus of the amygdala (CNA) for its acquisition (Gallagher et al., 1990; Hatfield et al., 1996; Han et al., 1997; Gallagher, 2000). However, an intact CNA is *not* required for the expression of a COR once it has been acquired (Hatfield et al., 1996; Groshek et al., 2005). Note also that both the original unconditioned orienting behavior to the CS and the habituation that occurs under repeated exposure to the CS remain intact after CNA lesions (Gallagher et al., 1990; Holland et al., 2002). Thus, the CNA is not directly involved in the representation of the orienting behavior per se, just in its conditioned acquisition. All CORs that have been examined to date depend on an intact dorsal striatum for their acquisition and/or expression (Hatfield et al., 1996; Han et al., 1997; Cardinal et al., 2002a), making the important link between the timing of CS-onset DA signals to the likely location of its learning effect.

3.2. Autoshaping

Autoshaping is the set of phenomena whereby subjects acquire new, sometimes rather bizarre, behaviors that involve orientation and approach towards, and often some sort of manipulation of, the CS in question, rather than the US. It was first described for pigeons (Brown and Jenkins, 1968). For example pigeons trained by pairing a localizable light with food delivery will approach and peck the light, even though its behavior has no contingency whatsoever with delivery of the reward and the location of the light CS is displaced from the location of the food source (US). Autoshaping has now also been shown in most other species studied (e.g., Holland, 1984; Locurto, 1981); in particular, it has been extensively studied in rodents (e.g., Cardinal et al., 2002b). Autoshaping CRs are also quite heterogeneous morphologically, and depend on several factors including the species involved, the sensory modality and the specific physical features of the CS, and others. A particularly critical parameter is the relative localizability of the CS. Thus, a localizable light more easily produces an autoshaping CR than a diffuse, non-localizable one (e.g., a house light), which almost never does. The nature of the US can also exert some modulatory effects on the morphology of the CR, but this is typically less influential than the features of the CS itself (Holland, 1977, 1984; Terrace, 1981).

The similarity of autoshaping to the COR described above is apparent because autoshaping also depends on the presence of an unconditioned response to the CS that is then reinforced through conditioning. For example, pigeons exhibit a low spontaneous rate of pre-pecking, or even outright pecking, directed towards a localized light source prior to training (Terrace, 1981; Jenkins, 1973; Gibbon et al., 1975). As Wasserman (1981) describes: "An orderly sequence of behavior, namely, orientations, approaches, and key-directed pecks, often precedes the first recorded keypeck (e.g., Brown and Jenkins, 1968; Rachlin, 1969; Wessells, 1974)." Note that orienting behavior is actually a first component of the fully formed autoshaping response. Also, the unconditioned response to the CS can be fairly latent in some cases, only fully emerging after repeated rewarded learning trials.

As is the case for the COR, the CNA is important for the acquisition, but not expression, of autoshaping CRs (e.g., Killcross et al., 1997; Cardinal et al., 2002b). And, an intact ventral striatum is required for the acquisition and/or expression of autoshaping CRs (e.g., Cardinal et al., 2002b), again tying CS driven DA to its likely locus of action.

3.3. Training of CS-generated CRs through phasic DA

For both the COR and autoshaping cases, the PVLV model suggests that phasic DA bursts driven by the CNA (LV_e in the PVLV model) at CS onset serve to reinforce the existing or latent CR behaviors. Because these behaviors are mediated by neural connections that exist prior to conditioning, it follows that a phasic dopamine signal can serve to strengthen these connections, resulting in the conditioned form of these responses. The U-shaped curve associated with COR learning is exactly what would be predicted given that the CS-associated DA burst itself takes some time to become established: early in training, the absence of this DA burst allows the OR to become habituated, and then it is reaquired as the DA burst grows.

The timing of the CS phasic DA burst should be appropriate for these CS-generated behaviors, as contrasted with other US-generated behaviors, which would be reinforced by US-mediated dopamine bursts. This is consistent with the fact that CNA lesions do not affect the acquisition of US-generated CRs (e.g., Han et al., 1997). Fig. 4 illustrates how the acquisition of the two categories of striatum-dependent CRs can be dissociated by CNA lesions, based on the premise that US-generated CRs are trained by US-onset triggered dopamine bursts and CS-generated CRs are trained by CS-onset dopamine bursts.

As noted, both autoshaping and COR are dependent on an intact striatum for their acquisition, implicating plasticity in this area. That phasic dopamine signal in the striatum may be directly responsible for this plasticity is supported by the some relevant pharmacological data. Highly suggestive, if not definitive, evidence in support of this hypothesis is that D_1 receptor antagonists have been shown to interfere with the acquisition of versions of the striatum dependent CRs described here (Cardinal et al., 2002a; Parkinson et al., 1998; Day and Carelli, 2007; Di Ciano et al., 2001; Parkinson et al., 2002). We say this cannot be considered definitive because it is not yet known if these D_1 antagonists are interfering with the effects of *phasic* DA signals, tonically derived extracelluar dopamine levels, or both.

Importantly, the striatum apparently does not tend to show sustained firing to stimuli (Apicella et al., 1992; Cromwell and

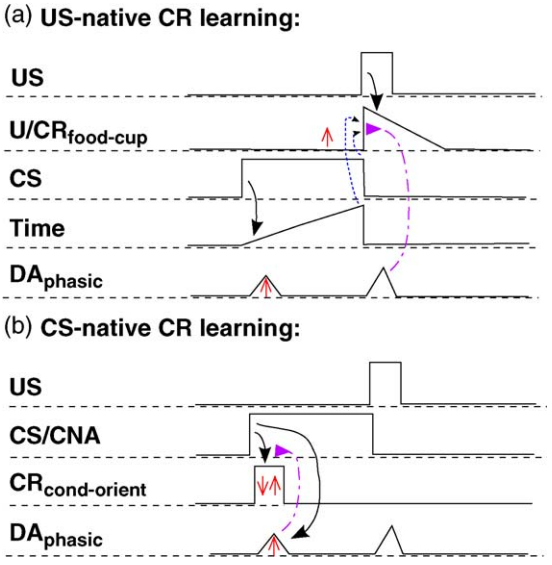

Fig. 4. US-generated vs. CS-generated CR learning. Complex striatum dependent conditioned responses can be subdivided into those that are innately triggered by the reward stimulus (US) (e.g., food-cup behavior), versus those innately triggered by the neutral stimulus (CS) (e.g., conditioned orienting, autoshaping), both prior to any conditioning. According to PVLV, these two categories are trained differentially by the two different types of phasic DA burst—US-onset and CS-onset, as shown here. (a) US-generated CR learning: the phasic DA burst triggered by the US can strengthen food-cup behavior (purple arrow). (b) CS-generated CR learning: the acquired phasic DA burst to CS onset is able to strengthen the existing representation of the orienting response, enhancing the behavior. CNA lesions selectively eliminate this category of CR, suggesting that the CNA plays a role very much like LV_e as proposed in the PVLV algorithm.

Schultz, 2003), meaning that a precisely timed phasic DA burst at CS-onset would seem to be particularly critical for strengthening the weights that drive approach-to-CS behavior: behaviors that are not natively activated by CS onset should not be reinforced by this CNA-mediated phasic dopamine signal.

One important caveat for interpreting the effects of CNA lesions is that the CNA also represents the gateway from the amygdala to a number of primitive brainstem nuclei (e.g., Amaral et al., 1992), and has been shown to be important for conditioning effects involving more basic autonomic brainstem-level functions (e.g., conditioned insulin release; Roozendaal et al., 1990). Although these may have some influence on higher level functions, we do not think they are likely to mediate the conditioning effects described above.

To summarize, the PVLV model predicts that all CRs natively associated with a CS and dependent upon phasic dopamine for acquisition (e.g., by virtue of depending on the striatum, which is known to be strongly modulated by phasic dopamine) should be affected by CNA lesions. To our knowledge, this is the case wherever relevant data is available. In the following sections, we review other related findings that match predictions that follow from the PVLV model.

3.4. CS-generated CRs are not susceptible to the blocking effect

Unlike most other forms of conditioned learning, CS-generated CRs are not susceptible to the well-known blocking effect (Kamin, 1969a,b), where an existing CS–US association is capable of blocking the acquisition of a second CS–US association, when both CSs are presented at the same time. For example, Tomie (1981) trained pigeons in an autoshaping procedure using a visual CS consisting of localizable green light (CS_a). Next, the subjects received a second training session in which a vertical white line (CS_b) was superimposed upon the green light, creating a green + white-line compound stimulus. Finally, the animals were tested on just the vertical white line CS_b. Remarkably, the animals exhibited high rates of pecking to the CS_b alone on the very first test trials. Tomie (1981) performed four separate variations of this same basic experiment and all four variations produced highly consistent results showing a complete lack of blocking (Tomie, 1981).

This striking exception to the blocking effect would seem potentially challenging to the PVLV model, which is based on the Rescorla–Wagner rule that exhibits the blocking effect. However, the blocking effect in PVLV is exhibited only by the US-associated PV system, whereas the CS-associated LV system only experiences a partial (and very slowly emerging) reduction in CS-generated DA firing due to the LV_i system as described earlier. Thus, we would predict that CS-associated CRs would not be affected by blocking manipulations, while US-associated ones would, consistent with the Tomie (1981) data. Specifically, during the compound training phase, the green light is still present to drive DA firing that does not get predicted away. And, this phasic DA burst is perfectly timed for training up an association between the new CS_b and the pecking behavior.

One further prediction of the PVLV model in this regard is that extensive levels of prior conditioning should reduce the relative rate of CS_b acquisition in such a paradigm, a partial blocking effect, due to the slow reduction (but not elimination) of DA bursting, as discussed earlier.

3.5. CNA lesions impair generalized PIT

In addition to the kinds of explicitly defined conditioned responses already discussed, there is an additional set of effects produced by Pavlovian paradigms called Pavlovian instrumental transfer (PIT) (e.g., Cardinal et al., 2002a). This refers to an enhancement in instrumental responding, e.g., an increase in both the number and vigor of bar presses, when animals are tested in the presence of the CS, relative to the CS being absent. It is now well established that there are two clearly dissociable forms of PIT: (1) *general* (genPIT), and (2) *outcome-specific* (osPIT). The latter is restricted to CSs paired with the same US used in the test instrumental condition, while the former involves CSs paired with any US (Cardinal et al., 2002a; Corbit and Balleine, 2005). Note that, for the case of osPIT, the net increase in responding will actually be a combined effect of genPIT + osPIT, since a US-specific CS is also "any" CS (Corbit and Balleine, 2005).

An intact CNA is critical for the acquisition and/or expression of genPIT, but not for osPIT (Corbit and Balleine, 2005). This effect on genPIT is generally consistent with the biological interpretation of PVLV described here in that the CNA (LV_e), having become activated during the CS, drives phasic dopamine firing in a *nonlocalized* way. This nonlocalized activity, in turn, can facilitate any instrumental responding in a nonspecific way through the generally excitatory effect of dopamine on striatal Go pathway MSNs (via D_1Rs) and concomitant inhibitory effect on striatal NoGo MSNs (D_2Rs) (Frank, 2005; Houk et al., 1995; Mink, 1996).

While initially attractive, a simple account of genPIT in terms of CS-onset driven phasic dopamine producing generalized activation is probably too simplistic. The reason is that the genPIT effect is observed throughout the duration of CS occurrence, and not just at the phasic onset of the CS. While increased responding for a short time period (e.g., < 100 ms) immediately after CS-onset might be expected from phasic DA firing alone, extracellular dopamine levels quickly return to prior ambient levels after the phasic burst has stopped (Floresco et al., 2003), but subjects continue to exhibit genPIT effects throughout CS duration.

A possible answer to this puzzle involves the ability of the system to modulate tonic dopamine firing, which can have the requisite longer-lasting effects. Tonic DA activity is itself under the control of the ventral pallidum (VP), the major output nucleus of the ventral striatum. Increased VP activity serves to depress tonic DA firing, while decreased VP activity increases it (Floresco et al., 2003). Interestingly, this effect appears not to be via changes in tonic firing rates of individual DA cells per se, which remain highly regular at 2–5 Hz, is instead due to the toggling of DA cells from a quiescent to tonically active state and vice versa. Furthermore, the level of extracellular DA in the striatum itself behaves like a Pavlovian conditioned response (CR), which is triggered by CSs associated with USs. Thus, we suggest that genPIT could result from the CNA-driven phasic DA at CS onset reinforcing the tonic dopamine CR, much as it reinforces orienting and autoshaping CRs. This would involve a similar phasic-DA modulated learning process in the subset of striatal MSN's that control the firing of neurons in the VP.

4. PVLV algorithm updates

We now turn to describing recent updates to the PVLV model, which, while important, do not substantially change the basic computational logic of the algorithm. Nor do they change the basic biological interpretation.

4.1. Novelty value (NV)

The most significant change has been the addition of a third component for driving the dopaminergic system to respond not only to reward prediction errors but also to pure novelty (e.g., Lisman and Grace, 2005). Fig. 5 shows a current biological interpretation of PVLV with the new *novelty value* (NV) component included. NV drives a phasic dopamine burst at the onset of any novel stimulus, with subsequent occurrences of the same stimulus

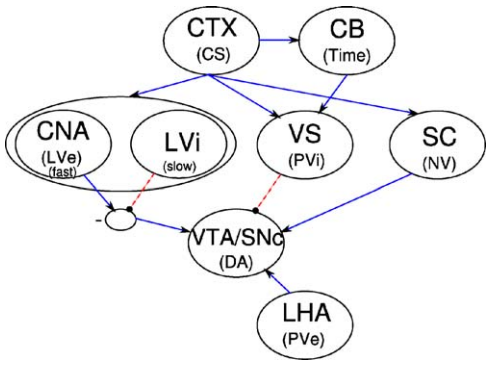

Fig. 5. Updated biological interpretation of PVLV including novelty value (NV) component novelty value (for visual stimuli) is deemed to be driven by known direct projections from the superior collicus (SC) to midbrain dopaminergic cells (VTA/SNc). *Legend:* CNA = central nucleus of the amygdala; VTA = ventral tegmental area; SNc = substantia nigra pars compacta (Dopamine areas); LHA = lateral hypothalamus; VS = ventral striatum (patch-like neurons); SC = superior colliculus; CB = cerebellum; CTX = Neocortex (sensory areas).

triggering progressively less dopamine according to a learned decay mechanism. For visual stimuli, the NV component is interpreted biologically in terms of known direct projections from the superior colliculus (SC) to midbrain dopaminergic cells (VTA/SNc) that have been shown to trigger DA cell firing that attenuates with repeated exposure (Comoli et al., 2003; Dommett et al., 2005; Humphries et al., 2006; Redgrave et al., 1999), or for hippocampal detection of novel stimulus features (Lisman and Grace, 2005).

This pattern is consistent with empirical data showing that DA cells respond phasically to the occurrence of novel stimuli (Horvitz, 2000; Kakade and Dayan, 2002). A recent study showed that preferences for novelty-based choice are accompanied by striatal activations modeled by adding a novelty term to the value of each action, such that reward value and novelty contribute additively to action selection (Wittmann et al., 2008).

The functional role of NV-derived burst firing in PVLV is similar to that proposed by Kakade and Dayan (2002), encouraging a kind of computational exploration with regard to the maintenance of new stimuli which might turn out to be rewarding. In the context of working memory this implies that potential CSs are more likely to be updated and maintained in working memory when they are novel, thereby speeding the discovery of appropriate working memory representations that can contribute to favorable performance. In addition to this simple 'novelty bonus,' other mechanisms, likely dependent on more elaborate prefrontal-dependent processes, may contribute to exploration based on *uncertainty* about whether a given action might lead to better outcomes than have been experienced in the past (Dayan and Sejnowski, 1996; Frank et al., 2009).

4.2. $\dot{Y}$ mechanism for DA bursting

The second update is in the way that DA activation is computed to produce the burst pattern of firing observed empirically. That is, DA cells do not persistently burst fire during the entire duration of a sustained CS or US, but rather exhibit only a transient phasic burst. In the original PVLV model, the transient nature of DA bursting was accomplished by a synaptic depression mechanism causing LV neurons to stop firing continuously to a sustained CS stimulus. However, it turned out that this mechanism performed suboptimally in discriminating among different potential CSs. For this reason, and because CNA neurons, thought to be the neural substrate for the LV_e representations, are known to fire in a sustained manner (Ono et al., 1995), we replaced the synaptic depression mechanism with a mechanism downstream from the LV_e representations themselves. The available empirical data suggests that the basis for the short bursting pattern is most likely within the midbrain DA nuclei themselves, in the pedunculopontine tegmental (PPT) nucleus, or a combination of both.

Therefore, we replaced the synaptic depression mechanism with a more straightforward temporal derivative $\dot{Y}$ (pronounced "Y-dot") approach (Sutton and Barto, 1990). Note that this does not make PVLV significantly more similar to TD, because the temporal derivative is not used for learning within the PVLV system in the same way it is used within TD—it is merely used to drive appropriate burst firing in the DA outputs.

4.3. Updated computational implementation

Next, we summarize the way that the new NV system contributes to DA firing and then how it is combined with PV and LV. First, to mirror the known attenuation of superior colliculus-driven DA firing that occurs with repeated exposure to novel visual stimuli (Comoli et al., 2003; Dommett et al., 2005; Humphries et al., 2006; Redgrave et al., 1999), the NV system learns to decrease the novelty value as a functon of each encounter with the relevant stimulus:

$$\Delta w_i^t = -\epsilon NV^t x_i^t \tag{1}$$

where NV^t is the current novelty value at time step t (initialized to 1, and decays slowly to 0 through this learning mechanism with a low learning rate ϵ; Δw_i^t is change in weight from sending unit i with activation value x_i^t).

The dopamine system (VTA/SNc) integrates each of the now three inputs (PV, LV, NV), using a temporal derivative computation ($\dot{Y}$) to produce brief bursts or dips relative to a baseline level of activation. As emphasized in the original PVLV paper, the key issue is *when* to use each of the above values: If primary rewards are present or expected by PV, then the PV system dominates contributions to DA; otherwise, LV + NV can drive DA activation. There is a simple threshold applied to the PV_i activity to determine if there is a significant (positive or negative) reward expectation (with .5 considered to be the baseline neutral value, 0 a strong negative input, and 1 a strong positive one):

$$PV_{filter} = PV_e \text{ value or } (PV_i > .8 \text{ or } PV_i < .2) \tag{2}$$

Critically, this condition is also used to determine when to train the LV system, where LV_e in the model is trained to approach the PV_e (US) values. Biologically, we think that the LV system is actually being trained by a combination of direct PV_e (US) signals from the LHA and phasic dopamine signals at the time of US-onset. Importantly, we hypothesize that CS-driven dopamine signals, which arise from the action of the LV system itself, arrive too late to drive useful synaptic changes in the LV system (cf., Tan and Bullock, 2008), as discussed in detail earlier.

The full DA equation with temporal derivatives (t = current trial, $t-1$ is previous trial) is:

$$\delta = \begin{cases} (PV_\delta^t - PV_\delta^{t-1}), & \text{if } PV_{filter} \\ (LV_\delta^t - LV_\delta^{t-1}) + (NV^t - NV^{t-1}), & \text{otherwise.} \end{cases} \tag{3}$$

The delta-rule/Rescorla–Wagner learning equation for the PV system is:

$$\Delta w_i^t = \epsilon(PV_e^t - PV_i^t)x_i^t \tag{4}$$

where Δw_i^t is the change in weight at time step t, from sending unit i with activation x_i^t. This is computed at every time step. In contrast,

the LV learning rule is filtered by the PV filter:

$$\Delta w_i^t = \begin{cases} \epsilon(\mathrm{PV}_e^t - \mathrm{LV}_e^t)x_i^t, & \text{if } \mathrm{PV}_{filter} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Note that the PV, LV, and NV values are all represented as scalar values using coarse-coded neuronal representations, with the tuning curve of each unit shifted to represent a different value (here, three units most strongly represent values of 0, .5, and 1). The initial bias for all the values is to represent a .5, except NV which is initialized with a value of 1.0 that then decreases with experience as described above.

Finally, to further clarify the role of the PV_{filter}, it serves to differentiate between two categories of timesteps: those timesteps when a US occurs (or would have been expected to occur), versus all other timesteps. CS occurrences, both when novel and after training, typically fall into the latter category. Thus, as they are gradually acquired, LV_e signals essentially 'take over' from the NV signals (as they wane) in driving phasic DA firing at CS-onset.

5. Comparison between PVLV and TD

In this section, we will expand upon our introductory discussion comparing PVLV to TD in several critical respects.

TD can be thought of as a delta-rule learning mechanism that straddles two adjacent points in time, with the present serving as a training signal for the immediate past. In contrast, the Rescorla–Wagner version of the delta rule operates within a single time period during which both an expectation and US outcome are encoded (active) and compared. Thus, at a mathematical level, TD and PVLV (which can be thought of as an elaborated version of Rescorla–Wagner) share the basic delta-rule dynamic, but they differ principally in how they use time for computing the DA delta. This difference can be captured to some extent within the TD framework itself, through the use of a time-smearing mechanism ("eligibility traces") in the TD(λ) formulation (Sutton and Barto, 1998). The λ parameter controls the exponential rate of decay of prior stimulus representations which are available to the learning mechanism, with $\lambda = 0$ having no such trace at all, and $\lambda = 1$ having an infinite trace. For TD($\lambda = 0$), i.e., no temporal smearing, each moment in time is completely distinct, every individual timestep is indispensible, and temporal chaining is essential for bridging over delays. On the other hand, as one approaches the other extreme of TD($\lambda = 1$), chaining becomes much less important, and the system converges in some ways back to the simpler Rescorla–Wagner atemporal behavior.

A longstanding empirical issue for TD($\lambda = 0$) has been that phasic dopamine firing ought to be seen chaining backward in a 'bucket brigade-like' manner during the CS–US interstimulus interval (ISI) early in training. This is because the way that TD works is to pass the prediction error signal backward one timestep at a time with each training trial, implying that phasic DA firing ought to be seen between the CS-onset and the US. This behavior has generally not been observed empirically (but see Niv et al. (2005) for an argument that evidence for chaining may be buried in the noisy ISI period). Instead, firing is generally interpreted as jumping directly to the time of CS-onset by most authors (e.g., Fiorillo et al., 2005; Pan et al., 2005). Specifically addressing this issue, Pan et al. (2005) recently presented new dopamine firing data that largely replicated earlier patterns of firing, and then showed that only a TD(λ) model, but not TD(0), was able to reasonably simulate the empirical data. In particular, large λ values (close to 1) were required, such that learning from one time step almost completely generalizes to all previous timesteps, thereby almost reducing TD to just the standard Rescorla–Wagner rule.

Even with the addition of eligibility traces to the TD(λ) framework, however, other significant issues for TD as a biological model remain. First, TD(λ) does not make a very clear distinction between CSs that persist through to the time of reward (delay conditioning), and those that do not (trace conditioning): the exponential trace enables similar learning to take place in either case. However, it is well established that trace conditioning depends on the integrity of both the hippocampus and prefrontal cortex, whereas delay paradigms do not (e.g., Weible et al., 2000; Kronforst-Collins and Disterhoft, 1998). The standard account of this dissociation is that these additional brain systems are necessary for actively maintaining a neuronal representation of the stimulus through to the point of reward, bridging the gap so that an association can be established. It is unclear why these separate memory systems would be required within the unitary TD(λ) framework, especially because a large λ is generally required to account for even the delay conditioning data, so it would not make sense to attribute the exponential trace function to the hippocampus and prefrontal cortex.

In contrast, PVLV makes a clear distinction between trace and delay paradigms, because it can only learn about a CS at the time of the US. If the CS is no longer present, some other internal representation of it must be preserved to bridge the associative gap. Importantly, a mechanism for updating and maintaining working memory representations is learnable as a function of task demands (O'Reilly and Frank, 2006; Hazy et al., 2006, 2007). This learned working memory system has the advantage of being able to hold onto information for varying amounts of time depending on task demands, whereas the λ value of the eligibility trace is generally a fixed parameter, which may or may not be appropriate for a given task.

More generally, the unitary nature of the TD framework does not seem compatible with the relatively large and diverse cast of brain areas involved in driving dopamine firing. In contrast to TD, PVLV predicts that it should be possible to doubly-dissociate the CS and US associated DA firing behavior empirically. Specifically, we would predict that lesions/inactivation of the medial segment of the CNA should prevent the acquisition, and expression, of DA bursting to CS-onset, while lesions of the ventral striatum ought to prevent the loss of DA bursting at US-onset (even as new firing to CS-onset is acquired normally).

Finally, the restriction that the LV system cannot reinforce itself, which has clear computational motivations as described above, gives rise to another important difference between TD and PVLV in the context of second and higher order conditioning. Because a CS-elicited DA burst would not be expected to train further CS-CS associations within the PVLV system, this mechanism should not support higher order conditioning. In contrast, the unitary nature of TD causes it to automatically and easily support arbitrarily high orders of conditioning. As we pointed out in the original paper (O'Reilly et al., 2007), there is a dearth of evidence for higher order conditioning beyond second order conditioning (Denny and Ratner, 1970; Dinsmoor, 2001). While the absence of direct evidence is clearly not evidence for absence, we suggest that even second order conditioning displays characteristics quite distinct from first order, in a pattern suggesting it may *not* be primarily dependent on phasic DA firing (e.g., Hatfield et al., 1996).

In summary, both PVLV and TD(λ) models (with λ closer to 1 than 0) make a lot of the same predictions that are consistent with known empirical phenomena, because both use a similar treatment of time, and both are fundamentally delta-rule/Rescorla–Wagner based mechanisms. However, PVLV also makes other predictions that TD does not make, which also seem consistent with available data. Some researchers will likely prefer the theoretical elegance of the TD framework, particularly as a normative model in the context of traditional artificial intelligence

(AI; e.g., for solving grid-world problems that require learning to reinforce actions that lead to the greatest long-term reward). And, surely various additional modifications can continue to be made to make it conform to the relevant data. However, we argue that the striking alignment between the idiosyncratic features of the PVLV mechanism and the available empirical data would suggest that PVLV may provide a better biological level model. Thus, we would argue that, as a biological model, it may be more parsimonious to abandon a temporal chaining framework of TD entirely and embrace the Rescorla–Wagner picture directly, instead of so dramatically blurring the temporal chaining element of TD by using large λ values.

6. Discussion

We have presented a mechanistically explicit model of phasic dopamine firing and associated Pavlovian learning phenomena, which was motivated initially by computational considerations, and appears to be consistent with a wide range of biological and behavioral data as reviewed above. We have extended the original model (O'Reilly et al., 2007) in a couple of straightforward ways (adding a novelty-driven phasic DA firing mechanism, and a better mechanism for detecting and learning about the omission of expected rewards), and modified it to better fit with the known biological mechanisms of phasic DA burst generation (though many biological details remain to be resolved regarding this mechanism). As we continue to explore the biological mappings and implications of the PVLV framework, a growing body of literature appears to be consistent with the specific divisions of labor (PV vs. LV) that it posits. Although the temporal differences (TD) framework is arguably more mathematically elegant, and powerful in many respects, we consider the rich mapping that PVLV provides onto the underlying biology to be a considerable strength.

In the next section, we elaborate a number of testable predictions that the PVLV framework makes, in the hopes of stimulating empirical research that might serve to further test the core ideas behind the PVLV model. After that, we discuss a set of other brain areas outside the PVLV system and their potential role in various important conditioning phenomena—further development of explicit computational models of these brain areas will be important for clarifying the exact contributions that PVLV makes to overall learning behavior, and is a major goal of ongoing research.

6.1. Key implications and predictions of the PVLV framework

The six biological hypotheses behind the PVLV model enumerated above suggest several critical predictions that might be used to confirm or invalidate specific aspects of the model. Predictions that we consider most important are listed below.

1. The hypothesis that the LHA is the main excitatory driver for phasic DA burst firing (PV_e) implies that appropriate electrophysiological stimulation of the LHA should induce phasic DA bursting in the VTA and SNc, with appropriate latencies. Also, lesions of corresponding areas should eliminate/attenuate US-driven DA burst firing. This, in turn, should serve to eliminate/attenuate all of the acquired manifestations of Pavlovian conditioning that are thought to be dependent on a US-associated phasic DA signal, e.g., food cup behaviors. This prediction is largely noncontroversial and would probably be shared by most other models.
2. The hypothesis that patch-like cells of the ventral striatum play a role like that of PVLVs PV_i has several important implications: first, appropriate lesions of the ventral striatum, especially the NAc shell since it appears to be heavily patch-like, should prevent the "shunting" of phasic bursting for predicted rewards. This prediction is more or less shared with the model(s) of Brown et al. (1999) and Tan and Bullock (2008), but contrasts with the that of Miller (2000), which places this functionality in the amygdala. Houk et al. (1995) placed this functionality in the dorsal striatum, but is generally similar in conception.
3. In addition, if the VS also contributes to the habenular mechanism described, VS lesions should also prevent the phasic dips seen with reward (US) omission, a prediction at least partially shared with Brown et al. (1999) and Tan and Bullock (2008), along with the next two.
4. A subpopulation of histologically identifiable patch-like cells of the VS, again especially in the NAc shell, should turn out to have direct, monosynapatic GABA-ergic connections onto the preponderance of midbrain DA cells of the VTA and SNc. Recall that such direct connectivity has been shown for striosomes of the dorsal striatum, but not yet for ventral (Joel and Weiner, 2000).
5. Either the ramping cells identified by Schultz and co-workers (Schultz et al., 1993b; Apicella et al., 1992; Cromwell and Schultz, 2003), and others (e.g., Deadwyler et al., 2004), or the cells seen firing only immediately before reward (Deadwyler et al., 2004), or perhaps both types, should turn out to be histologically identifiable as patch-like cells, e.g, they should be CB(−).
6. The hypothesis that PVLVs $LV_e \approx$ CNA makes the straightforward prediction that mCNA inactivation should prevent the acquisition of phasic DA burst firing at CS-onset. This is one of the most central implications of the PVLV framework. This prediction is shared with the Miller (2000) model, but contrasts with the Brown et al. (1999) and Tan and Bullock (2008) model(s), as well as Houk et al. (1995), which places this functionality in the striatum. Perhaps most importantly, the dissociation between this prediction and the four prior ones involving the striatum is an important source of differentiation for PVLV relative to most other models, including TD-based models, since these all tend to point to a single mechanism and/or anatomical location underlying both excitation and inhibition.
7. With regard to mechanisms that prevent self-training of the CNA (LV_e), many empirical tests are possible. For example, if phasic DA signals are important for training CS–US pairing for CNA neurons, then DA antagonists injected into the CNA should prevent the acquisition of CS-associated DA firing. If this is what is found, then we would predict that the timing dependence should be such that CS activation to the CNA must be sustained for a significant interval prior to the arrival of a DA signal for synaptic changes to take place, an interval longer than that required for the autofeedback loop from CNA to VTA and back. This idea is also consistent with the existing behavioral data showing that a CS must generally precede a US by some minimal amount of time (e.g., at least 200 ms or more) for conditioning to occur (Schmajuk, 1997; Schneiderman, 1966; Smith, 1968). Depending on the results of such studies, one could also inject GABA antagonists in the latCNA to see if that interfered with a possible DA blocking mechanism as described earlier. This set of predictions is unique to PVLV.

In addition to suggesting specific predictions such as those just listed, the PVLV framework also raises many additional questions, the answers to which could eventually help to better characterize the overall workings of the dopamine system.

- If it is confirmed that the habenula is responsible for pausing tonic DA firing when reward is omitted, how does it actually know that a reward had been expected in the first place? Is it a signal from the VS, as we suggest? And, how does it know that

the reward did *not* occur? Is that a signal from the lateral hypothalamus, as we suggest?

- The early occurrence of a reward prevents a subsequent dip at the normal (expected) time (Fig. 1d; e.g., Schultz and Dickinson, 2000). This implies some sort of smart resetting process—how might this work? We would predict that it would involve a turning off of the patch-like cells in the ventral striatum, which correspond to the PV_i mechanism. More generally, what kinds of resetting takes place once a primary reward (US) is received?
- The ramping cells in the VS (Schultz et al., 1993b; Cromwell and Schultz, 2003; Deadwyler et al., 2004) seem to be integrating timing information in order to peak at the time of anticipated US. The source of this evolving time representation remains an empirical question. One obvious candidate is the cerebellum, widely thought to be important for representing time (Mauk and Buonomano, 2004). This would require cerebellar input to influence VS neurons, which does not happen through direct projections, but there are indirect pathways via cerebellothalamic and thalamostriatal projections (Bentivoglio et al., 1988); and via the frontal cortex (Lustig et al., 2005). Another proposal is that striosomes can actually exhibit timing dynamics themselves (Brown et al., 1999).
- To examine the nature of the LV_i-like signal that reduces CS firing with extensive training, one could change the context after extensive training, and see if that restores the original high level of CS firing. This would suggest that the LV_i mechanism reflects contextual learning, consistent with the other data showing that CS onset bursts can be blocked completely by a prior predictive CS.
- Can a true second order (i.e., without US reinforcement of the second-order CS) acquired phasic DA burst be demonstrated? So far, the empirical data seems unclear on this issue. Hatfield et al. (1996)'s findings with second order COR suggest a second order phasic DA burst might be possible, but the well-recognized difficulty in obtaining Pavlovian conditioning higher than second suggests that phasic DA bursts beyond second ought not to be seen, or least extremely difficult to obtain.
- How is punishment represented in the neural systems associated with the PVLV framework? In PVLV, we use .5 as a neutral value (corresponding to the baseline DA tonic firing rate), with 1 being positive value and 0 being negative value relative to this .5 baseline. This corresponds to the idea that aversive stimuli or negative prediction errors result in dips/pauses in DA tonic firing, which has considerable empirical support (Mirenowicz and Schultz, 1996; Tobler et al., 2003; Ungless et al., 2004). However, others have argued that phasic bursts of DA accompany aversive inputs (e.g., Horvitz, 2000). Our preferred interpretation of this data is that it reflects a novelty signal of the form computed by the new NV system in PVLV. More definitive empirical tests of this issue would be quite valuable.

6.2. Beyond the core PVLV model: ancillary mechanisms and future development

The main focus of this article has been to provide a biological interpretation of the core PVLV model in accounting for the related empirical findings associated with phasic dopamine firing and Pavlovian conditioning. Accordingly, the discussion has been restricted to those phenomena for which PVLV may provide a reasonably complete account. However, phasic DA firing and Pavlovian conditioning are both obviously relevant in a much broader context of behavioral and cognitive phenomena including response selection, decision making, and executive function generally. In particular, they are both directly relevant to the related domain of instrumental conditioning. We have made substantial initial progress integrating PVLV with learning in the prefrontal cortex and basal ganglia, in the context of our PBWM model that learns to actively maintain and output information (O'Reilly and Frank, 2006; Hazy et al., 2006, 2007).

By far the biggest issue not yet encompassed within the PVLV/PBWM framework is the role of the basolateral complex of the amygdala, which has been preferentially implicated in a wide variety of learning-related phenomena including: second order Pavlovian conditioning; conditioned reinforcement (CRf); instrumental conditioning generally, and; incentive salience. Considering their colocalization with the same anatomical structure and their often similar patterns of activation (Ono et al., 1995), there is a rather striking pattern of *dissociation* that has been demonstrated between lesions of the BLA versus its amygdalar partner, the CNA, when it comes to behavioral effects. As we have argued in this paper, we think the CNAs role has largely to do with driving phasic DA firing at CS-onset, in addition to a number of other subcortical effects, of course.

On the other hand, the BLA seems to be critical for a whole myriad of phenomena that, by our analysis at least, seem not to be specifically dependent on phasic dopamine firing, at least not directly. This distinction seems to underly a functional division of labor between the BLA and the CNA vis-a-vis their respective roles in Pavlovian and instrumental conditioning. A major thrust of our future efforts to extend the PVLV framework will be to include a BLA component in our models so as to begin to encompass these additional phenomena, including instrumental conditioning, in the framework. Anatomically, the BLA sends significant input to the ventral striatum and the orbital prefrontal cortex, and this may enable it to provide much richer detailed signals about CS and US information, as contrasted with the global DA signal putatively produced by the CNA.

6.3. Conclusion

In this paper we have tried to show how a biological interpretation of the PVLV algorithm can begin to account for many of the important, if sometimes subtle, aspects of phasic DA firing. Specifically, we identified six specific hypotheses that collectively characterize a biological interpretation of PVLV and reviewed the empirical evidence relevant to each. While still far from a complete picture, we believe that the PVLV framework makes significant progress in a positive direction. We have identified a number of empirical predictions and suggested possible experiments that would help to test our key hypotheses and help to push the collective understanding further.

References

Ahn, S., Phillips, A.G., 2003. Independent modulation of basal and feeding-evoked dopamine efflux in the nucleus accumbens and medial prefrontal cortex by the central and basolateral amygdalar nuclei in the rat. Neuroscience 116, 295–305.

Amaral, D.G., Price, J.L., Pitkanen, A., Carmichael, S.T., 1992. Anatomical organization of the primate amygdaloid complex. In: The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction. Wiley-Liss, New York, pp. 1–66.

Andrzejewski, M.E., Spencer, R.C., Kelley, A.E., 2005. Instrumental learning, but not performance, requires dopamine d1-receptor activation in the amygdala. Neuroscience 135 (2), 335–345.

Apicella, P., Scarnati, E., Ljungberg, T., Schultz, W., 1992. Neuronal activity in monkey striatum related to the expectation of predictable environmental events. Journal of Neurophysiology 68, 945–960.

Aston-Jones, G., Smith, R.J., Moorman, D.E., Richardson, K.A., 2009. Role of lateral hypothalamic orexin neurons in reward processing and addiction. Neuropharmacology 56 (Suppl. 1), 112–121.

Bao, S., Chan, V.T., Zhang, L.I., Merzenich, M.M., 2003. Suppression of cortical representation through backward conditioning. Proceedings of the National Academy of Sciences of the United States of America 100, 1405–1408.

Baxter, M.G., Murray, E.A., 2002. The amygdala and reward. Nature Reviews Neuroscience 3, 563–572.

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

Bayer, H.M., Lau, B., Glimcher, P.W., 2007. Statistics of midbrain dopamine neuron spike trains in the awake primate. Journal of Neurophysiology 98 (3), 1428–1439.

Belova, M.A., Paton, J.J., Morrison, S.E., Salzman, C.D., 2007. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron 55 (6), 970–984.

Bentivoglio, M., Minciacchi, D., Morinari, M., Granato, A., Spreafico, R., Macchi, G., 1988. The intrinsic and extrinsic organization of the thalamic intralaminar nuclei. In: Bentivoglio, M., Spreafico, R. (Eds.), Cellular Thalamic Mechanisms. Excerpta Medica, Amsterdam, pp. 221–237.

Berendse, H.W., Groenewegen, H.J., Lohman, A.H., 1992. Compartmental distribution of ventral striatal neurons projecting to the mesencephalon in the rat. Journal of Neuroscience 12 (6), 2079–2103.

Keri, N., Nagy, S., Moustafa, H., Myers, A., C.E., Daw, N., Dibo, G., Takats, A., Bereczki, D., Gluck, M.A., 2009. Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young parkinson's patients. Brain 132 (9), 2385–2395.

Braver, T.S., Cohen, J.D., 2000. On the control of control: The role of dopamine in regulating prefrontal function and working memory. In: Monsell, S., Driver, J. (Eds.), Control of Cognitive Processes: Attention and Performance, vol. xviii. MIT Press, Cambridge, MA, pp. 713–737.

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

Brown, J., Bullock, D., Grossberg, S., 2004. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. Neural Networks 17, 471–510.

Brown, P., Jenkins, H.M., 1968. Auto-shaping of the pigeon's key-peck. Journal of the Experimental Analysis of Behavior 11, 1–8.

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

Cardinal, R.N., Parkinson, J.A., Lachenal, G., Halkerston, K.M., Rudarakanchana, N., Hall, J., Morrison, C.H., Howes, S.R., Robbins, T.W., Everitt, B.J., 2002b. Effects of selective excitotoxic lesions of the nucleus accumbens core, anterior cingulate cortex, and central nucleus of the amygdala on autoshaping performance in rats. Behavioral Neuroscience 116, 553–567.

Chong, V.Z., Skoblenick, K., Morin, F., Xu, Y., Mishra, R.K., 2006. Dopamine-d1 and -d2 receptors differentially regulate synapsin ii expression in the rat brain. Neuroscience 138 (2), 587–599.

Comoli, E., Coizet, V., Boyes, J., Bolam, J.P., Canteras, N.S., Quirk, R.H., Overton, P.G., Redgrave, P., 2003. A direct projection from superior colliculus to substantia nigra for detecting salient visual events. Nature Neuroscience 6, 974–980.

Cools, R., Altamirano, L., D'Esposito, M., 2006. Reversal learning in parkinson's disease depends on medication status and outcome valence. Neuropsychologia 44, 1663–1673.

Cools, R., Frank, M.J., Gibbs, S.E., Miyakawa, A., Jagust, W., D'Esposito, M., 2009. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 29 (5), 1538.

Corbit, L.H., Balleine, B.W., 2005. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. Journal of Neuroscience 25 (4), 962–970.

Cromwell, H.C., Schultz, W., 2003. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. Journal of Neurophysiology 89, 2823–2838.

Davis, M., Rainnie, D., Cassell, M., 1994. Neurotransmission in the rat amygdala related to fear and anxiety. Trends in neurosciences 17, 208–214.

Daw, N.D., Courville, A.C., Touretzky, D.S., 2003. Timing and partial observability in the dopamine system. In: Advances in Neural Information Processing Systems, vol. 15, MIT Press, Cambridge, MA.

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

Day, J.J., Carelli, R.M., 2007. The nucleus accumbens and pavlovian reward learning. The Neuroscientist 13 (2), 148–159.

Dayan, P., Niv, Y., 2008. Reinforcement learning: the good, the bad and the ugly. Current Opinion in Neurobiology 18 (2), 185–196.

Dayan, P., Sejnowski, T.J., 1996. Exploration bonuses and dual control. Machine Learning 25, 5–22.

Deadwyler, S.A., Hayashizaki, S., Cheer, J., Hampson, R.E., 2004. Reward, memory and substance abuse: functional neuronal circuits in the nucleus accumbens. Neuroscience and Biobehavioral Reviews 27, 703–711.

Denny, M.R., Ratner, S.C., 1970. Comparative Psychology: Research in Animal Behavior, revised edition. The Dorsey Press, Homewood, Illinois.

Di Ciano, P., Cardinal, R., Cowell, R.A., Little, S.J., Everitt, B.J., 2001. Differential involvement of nmda, ampa/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. Journal of Neuroscience 21 (2), 9471–9477.

Dinsmoor, J.A., 2001. Stimuli inevitably generated by behavior that avoids electric shock are inherently reinforcing. Journal of the Experimental Analysis of Behavior 75, 311–333.

Dommett, E., Coizet, V., Blaha, C.D., Martindale, J., Lefebvre, V., Walton, N., Mayhew, J.E., Overton, P.G., Redgrave, P., 2005. How visual stimuli activate dopaminergic neurons at short latency. Science 307, 1476–1479.

Doya, K., 2002. Metalearning and neuromodulation. Neural Networks 15, 495–506.

Dykman, R.A., 1965. Toward a theory of classical conditioning: cognitive, emotional, and motor components of the conditional reflex. In: Maher, B.A. (Ed.), Progress in Experimental Personality Research. Vol. 2. Elsevier Science & Technology Books, pp. 229–317.

Fallon, J.H., Ciofi, P., 1992. Distribution of monoamines within the amygdala. In: Aggleton, J.P. (Ed.), The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction. 1st ed. Wiley-Liss, New York, pp. 97–114.

Fiorillo, C.D., Tobler, P.N., Schultz, W., 2005. Evidence that the delay-period activity of dopamine neurons corresponds to reward uncertainty rather than back-propagating td errors. Behavioral and Brain Functions 1 (7), 1–5.

Fiorillo, C.D., Williams, J.T., 2000. Cholinergic inhibition of ventral midbrain dopamine neurons. Journal of Neuroscience 20 (20), 7855–7860.

Floresco, S.B., West, A.R., Ash, B., Moore, H., Grace, A.A., 2003. Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. Nature Neuroscience 6, 968–973.

Frank, M.J., 2005. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and non-medicated parkinsonism. Journal of Cognitive Neuroscience 17, 51–72.

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

Frank, M.J., Loughry, B., O'Reilly, R.C., 2001. Interactions between the frontal cortex and basal ganglia in working memory: a computational model. Cognitive, Affective, and Behavioral Neuroscience 1, 137–160.

Frank, M.J., O'Reilly, R.C., 2006. A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behavioral Neuroscience 120, 497–517.

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

Frey, U., Matthies, H., Reymann, K., Matthies, H., 1991. The effect of d1 dopaminergic receptor blockade during tetanization on the expression of long-term potentiation in the rat ca1 region in vitro. Neuroscience Letters 129, 111–114.

Frey, U., Schroeder, H., Matthies, H., 1990. Dopaminergic antagonists prevent long-term maintenance of posttetanic ltp in the ca1 region of rat hippocampal slices. Brain Research 552, 69–75.

Fudge, J.L., Haber, S.N., 2000. The central nucleus of the amygdala projection to dopamine subpopulations in primates. Neuroscience 97, 479–494.

Fukuda, M., Ono, T., 1993. Amygdala-hypothalamic control of feeding behavior in monkey: single cell responses before and after reversible blockade of temporal cortex or amygdala projections. Behavioural Brain Research 55 (2), 141–233.

Fuster, J.M., Uyeda, A.A., 1971. Reactivity of limbic neurons of the monkey to appetitive and aversive signals. Electroencephalography and Clinical Neurophysiology 30 (4), 281–293.

Gallagher, M., 2000. The amygdala and associative learning. In: Aggleton, J.P. (Ed.), The Amygdala: A Functional Approach. 2nd ed. Oxford University Press, Oxford, pp. 311–329.

Gallagher, M., Graham, P.W., Holland, P.C., 1990. The amygdala central nucleus and appetitive pavlovian conditioning: lesions impair one class of conditioned behavior. Journal of Neuroscience 14, 497–505.

Gallagher, M., Holland, P.C., 1993. Preserved configural learning and spatial learning impairment in rats with hippocampal damage. Hippocampus 2, 81–88.

Gallagher, M., Holland, P.C., 1994. The amygdala complex: multiple roles in associative learning and attention. Proceedings of the National Academy of Sciences of the United States of America 91, 11771–11776.

Gerfen, C.R., 1985. The neostriatal mosaic. I. compartmental organization of projections of the striatonigral system in the rat. Journal of Comparative Neurology 236, 454–476.

Gerfen, C.R., Herkenham, M., Thibault, J., 1987. The neostriatal mosaic. Ii. Patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. Journal of Neuroscience 7 (12), 3915–3934.

Gibbon, J., Locurto, C.M., Terrace, H.S., 1975. Signal-food contingency and signal frequency in a continuous trials auto-shaping paradigm. Animal Learning and Behavior 3, 317–324.

Glimcher, P.W., 2008. Perspectives from economics and neuroeconomics understanding risk: a guide for the perplexed. Cognitive Affective & Behavioral Neuroscience 8 (4), 348–354.

Grace, A.A., Bunney, B.S., 1979. Paradoxical gaba excitation of nigral dopaminergic cells: indirect mediation through reticulata inhibitory neurons. European Journal of Pharmacology 59 (3–4), 211–218.

Grace, A.A., Bunney, B.S., 1985. Opposing effects of striatonigral feedback pathways on midbrain dopamine cell activity. Brain Research 333 (2), 271–284.

Grace, A.A., Bunney, B.S., 2000. Electrophysiological properties of midbrain dopamine neurons. In: Floyd, M., Bloom, E., David, M., Kupfer, J. (Eds.), Psychopharmacology—The Fourth Generation of Progress. American College of Neuropsychopharmacology.

Graybiel, A.M., 1998. The basal ganglia and chunking of action repertoires. Neurobiology of Learning and Memory 70 (1–2), 119–136.

Groenewegen, H.J., Berendse, H.W., Wolters, J.G., Lohman, A.H., 1990. The anatomical relationship of the prefrontal cortex with the striatopallidal system the thalamus and the amygdala: evidence for a parallel organization. Progress in Brain Research 85, 95–116 (discussion 116–118).

Groshek, F., Kerfoot, E., McKenna, V., Polackwich, A.S., Gallagher, M., Holland, P.C., 2005. Amygdala central nucleus function is necessary for learning, but not expression, of conditioned auditory orienting. Behavioral neuroscience 119 (1), 202–212.

Han, J.S., McMahan, R.W., Holland, P., Gallagher, M., 1997. The role of an amygdalo-nigrostriatal pathway in associative learning. Journal of Neuroscience 17 (10), 3913–3919.
Hatfield, T., Han, J.S., Conley, M., Holland, P., 1996. Neurotoxic lesions of basolateral, but not central, amygdala interfere with pavlovian second-order conditioning and reinforcer devaluation effects. Journal of Neuroscience 16, 5256–5265.
Hazy, T.E., Frank, M.J., O'Reilly, R.C., 2006. Banishing the homunculus: making working memory work. Neuroscience 139, 105–118.
Hazy, T.E., Frank, M.J., O'Reilly, R.C., 2007. Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 362 (1), 105–118.
Hebb, D.O., 1949. The Organization of Behavior. Wiley, New York.
Heldt, S.A., Ressler, K.J., 2007. Lesions of the habenula produce stress- and dopamine-dependent alterations in prepulse inhibition and locomotion. Brain Research 1073 (4), 229–239.
Herkenham, M., Nauta, W.J.H., 1977. Efferent connections of the habenular nuclei in the rat. Journal of Comparative Neurology 187 (1), 19–47.
Holland, P.C., 1977. Conditioned stimulus as a determinant of the form of the pavlovian conditioned response. Journal of Experimental Psychology 3, 77–104.
Holland, P.C., 1984. Origins of behavior in pavlovian conditioning. In: The Psychology of Learning and Motivation, vol. 18, Academic Press, Inc., pp. 129–174.
Holland, P.C., Han, J.-S., Winfield, H.M., 2002. Operant and pavlovian control of visual stimulus orienting and food-related behaviors in rats with lesions of the amygdala central nucleus. Behavioral Neuroscience 116, 577–587.
Hollerman, J.R., Schultz, W., 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience 1, 304–309.
Horvitz, J.C., 2000. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience 96, 651.
Horvitz, J.C., Stewart, T., Jacobs, B.L., 1997. Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cats. Brain Research 759, 251–258.
Houk, J.C., Adams, J.L., Barto, A.G., 1995. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk, J.C., Davis, J.L., Beiser, D.G. (Eds.), Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, MA, pp. 233–248.
Huang, Y.-Y., Kandel, E., 1995. D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the ca1 region of the hippocampus. Proceedings of the National Academy of Sciences of the United States of America 92, 2446–2450.
Humphries, M.D., Stewart, R.D., Gurney, K.N., 2006. A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. Journal of Neuroscience 26 (50), 12921–12942.
Izhikevich, E.M., 2007. Solving the distal reward problem through linkage of stdp and dopamine signaling. Cerebral Cortex 17 (10), 2443–2452.
Jenkins, H.M., 1973. Effects of the stimulus reinforcer relation on selected and unselected responses. In: Hinde, A., Hinde, J.S. (Eds.), Constraints on Learning. Academic Press.
Ji, H., Shepard, P.D., 2007. Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a gaba-a receptor-mediated mechanism. Journal of Neuroscience 27 (26), 6923–6930.
Joel, D., Weiner, I., 2000. The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience 96, 451.
Kakade, S., Dayan, P., 2002. Dopamine: generalization and bonuses. Neural Networks 15, 549–559.
Kamin, L.J., 1969a. Predictability, surprise, attention and conditioning. In: Campbell, B.A., Church, R.M. (Eds.), Punishment and Aversive Behavior. Appleton-Century-Crofts, New York.
Kamin, L.J., 1969b. Selective association and conditioning. In: Mackintosh, N.J., Honig, W.K. (Eds.), Fundamental Issues in Instrumental Learning. Dalhousie University Press, Dalhousie, pp. 42–64.
Killcross, S., Robbins, T.W., Everitt, B.J., 1997. Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. Nature 388, 377.
Kronforst-Collins, M.A., Disterhoft, J.F., 1998. Lesions of the caudal area of rabbit medial prefrontal cortex impair trace eyeblink conditioning. Neurobiology of Learning and Memory 69, 147–162.
Lecourtier, L., DeFrancesco, A., Moghaddam, B., 2008. Differential tonic influence of lateral habenula on prefrontal cortex and nucleus accumbens dopamine release. European Journal of Neuroscience 27, 1755–1762.
LeDoux, J., 2003. The emotional brain, fear, and the amygdala. Cellular and Molecular Neurobiology 23 (4–5), 727–738.
Lee, H.J., Groshek, F., Petrovich, G.D., Cantalini, J.P., Gallagher, M., Holland, P.C., 2005. Role of amygdalo-nigral circuitry in conditioning of a visual stimulus paired with food. Journal of Neuroscience 25 (15), 3881–3888.
Legault, M., Wise, R.A., 2001. Novelty-evoked elevations of nucleus accumbens dopamine: dependence on impulse flow from the ventral subiculum and glutaminergic neurotransmission in the ventral tegmental area. European Journal of Neuroscience 13, 819–828.
Lisman, J.E., Grace, A.A., 2005. The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46, 703–713.
Ljungberg, T., Apicella, P., Schultz, W., 1992. Responses of monkey dopamine neurons during learning of behavioral reactions. Journal of Neurophysiology 67, 145–163.
Locurto, C.M., 1981. Contributions of autoshaping to the partitioning of conditioned behavior. In: Locurto, C.M., Terrace, H.S., Gibbon, J. (Eds.), Autoshaping and Conditioning Theory. Academic Press, New York, pp. 101–135.
Lustig, C., Matell, M.S., Meck, W.H., 2005. Not just a coincidence: frontal-striatal interactions in working memory and interval timing. Memory (Hove, England) 13, 441–448.
Matsumoto, M., Hikosaka, O., 2007. Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447, 1111–1115.
Matsumoto, O., Hikosaka, M., 2009. Representation of negative motivational value in the primate lateral habenula. Nature Neuroscience 12 (1), 77–84.
Mauk, M.D., Buonomano, D.V., 2004. The neural basis of temporal processing. Annual Review of Neuroscience 27 (1), 307–340.
McClure, S.M., Gilzenrat, M.S., Cohen, J.D., 2005. An exploration–exploitation model based on norepinepherine and dopamine activity. Advances in Neural Information Processing Systems 18 .
Miller, R., 2000. The amygdaloid complex: input processor for the midbrain dopaminergic nuclei and striatum. In: Miller, R., Wickens, J.R. (Eds.), Brain Dynamics and the Striatal Complex. Harwood Academic Publishers, Amsterdam, The Netherlands, pp. 77–110.
Mink, J.W., 1996. The basal ganglia: Focused selection and inhibition of competing motor programs. Progress in Neurobiology 50, 381–425.
Mirenowicz, J., Schultz, W., 1996. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature 379, 449.
Montague, P.R., Dayan, P., Sejnowski, T.J., 1997. A framework for mesencephalic dopamine systems based on predictive hebbian learning. The Journal of Neuroscience 16, 1936–1947.
Montague, P.R., Hyman, S.E., Cohen, J.D., 2004. Computational roles for dopamine in behavioural control. Nature 431 (7010), 760–767.
Mora, F., Rolls, E.T., Burton, M.J., 1976. Modulation during learning of the responses of neurons in the lateral hypothalamus to the sight of food. Experimental Neurology 53 (2), 508–519.
Moustafa, A.A., Cohen, M.X., Sherman, S.J., Frank, M.J., 2008. A role for dopamine in temporal decision making and reward maximization in parkinsonism. Journal of Neuroscience 28 (47), 12294–12304.
Nakamura, K., Ono, T., 1986. Lateral hypothalamus neuron involvement in integration of natural and artificial rewards and cue signals. Journal of Neurophysiology 55, 163–181.
Niv, Y., Duff, M.O., Dayan, P., 2005. Dopamine, uncertainty and td learning. Behavioral and Brain Functions 1 (6), 1–9.
Norgren, R., 1976. Taste pathways to hypothalamus and amygdala. Journal of Comparative Neurology 166 (1), 17–30.
Ono, T., Nakamura, K., Nishijo, H., Fukuda, M., 1986. Hypothalamic neuron involvement in integration of reward, aversion, and cue signals. Journal of Neurophysiology 56, 63–79.
Ono, T., Nishijo, H., Uwano, T., 1995. Amygdala role in conditioned associative learning. Progress in Neurobiology 46, 401–422.
Ono, T., Nishino, H., Sasaki, K., Fukuda, M., Muramoto, K.I., 1981. Monkey lateral hypothalamic neuron response to sight of food, and during bar press and ingestion. Neuroscience Letters 21 (1), 99–104.
O'Reilly, R.C., Frank, M.J., 2006. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Computation 18, 283–328.
O'Reilly, R.C., Frank, M.J., Hazy, T.E., Watz, B., 2007. Pvlv: The primary value and learned value pavlovian learning algorithm. Behavioral Neuroscience 121, 31–49.
O'Keefe, J., Bouma, H., 1969. Complex sensory properties of certain amygdala units in the freely moving cat. Experimental Neurology 23, 384–398.
Palminteri, S., Lebreton, M., Worbe, Y., Grabli, D., Hartmann, A., Pessiglione, M., 2009. Pharmacological modulation of subliminal learning in parkinson's and tourette's syndromes. Proceedings of the National Academy of Sciences of the United States of America 106 (45), 19179–19184.
Pan, W.-X., Schmidt, R., Wickens, J.R., Hyland, B.I., 2005. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. Journal of Neuroscience 25 (26), 6235–6242.
Parkinson, J.A., Dalley, J.W., Bamford, A., Fehrent, B., Robbins, T.W., Everitt, B.J., 1998. Effects of 6-ohda lesions of the rat nucleus accumbens on appetitive pavlovian conditioning. Journal of Psychopharmacology 12 (Suppl. A), A8.
Parkinson, J.A., Dalley, J.W., Cardinal, R.N., Bamford, A., Fehnert, B., Lachenal, G., Rudarakanchana, N., Halkerston, K.M., Robbins, T.W., Everitt, B.J., 2002. Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive pavlovian approach behaviour: Implications for mesoaccumbens dopamine function. Behavioural Brain Research 137, 149–163.
Paton, J.J., Belova, M.A., Morrison, S.E., Salzman, C.D., 2006. The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 439 (7078), 865–870.
Pavlov, I.P., 1927. Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. Oxford University Press, London.
Petrovich, G.D., Canteras, N.S., Swanson, L.W., 2001. Combinatorial amygdalar inputs to hippocampal domains and hypothalamic behavior systems. Brain Research Reviews 38 (1–2), 247–289.
Petrovich, G.D., Setlow, B., Holland, P.C., Gallagher, M., 2002. Amygdalo-hypothalamic circuit allows learned cues to override satiety and promote eating. Journal of Neuroscience 22 (19), 8748–8753.
Phillipson, O.T., 1978. Afferent projections to a10 dopaminergic neurones in the rat as shown by the retrograde transport of horseradish peroxidase. Neuroscience Letters 9, 353–359.

Pitkanen, A., 2000. Connectivity of the rat amygdaloid complex. In: Aggleton, J.P. (Ed.), The Amygdala: A Functional Approach. 2nd ed. Oxford University Press, Oxford, pp. 31–115.
Rachlin, H., 1969. Autoshaping of key pecking in pigeons wiht negative reinforcement. Journal of the Experimental Analysis of Behavior 12, 521–531.
Redgrave, P., Prescott, T.J., Gurney, K., 1999. Is the short-latency dopamine response too short to signal reward error? Trends in neurosciences 22, 146–151.
Rescorla, R.A., Wagner, A.R., 1972. A theory of pavlovian conditioning: variation in the effectiveness of reinforcement and non-reinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning. II. Theory and Research. Appleton-Century-Crofts, New York, pp. 64–99.
Roesch, M.R., Calu, D.J., Schoenbaum, G., 2007. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nature Neuroscience 10 (12), 1615–1624.
Roitman, M.F., Wheeler, R.A., Carelli, R.M., 2005. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. Neuron 45, 587–597.
Rolls, E.T., Burton, M.J., Mora, F., 1976. Hypothalamic neuronal responses associated with the sight of food. Brain Research 111 (1), 53–66.
Roozendaal, B., Oldenburger, W., Strubbe, J., Koolhaus, J., Bohus, B., 1990. The central amygdala is involved in the conditioned but not in the meal-induced cephalic insulin response in the rat. Neuroscience Letters 116, 210–215.
Rouillard, C., Freeman, A.S., 1995. Effects of electrical stimulation of the central nucleus of the amygdala on the in vivo electrophysiological activity of rat nigral dopaminergic neurons. Synapse (New York, NY) 21, 348–356.
Sanghera, M.K., Rolls, E., Roper-Hall, A., 1979. Visual responses of neurons in the dorsolateral amygdala of the alert monkey. Experimental Neurology 63, 610–626.
Santesso, D.L., Evins, A.E., Frank, M.J., Schetter, E.C., Bogdan, R., Pizzagalli, D., 2009. Single dose of a dopamine agonist impairs reinforcement learning in humans: evidence from event-related potentials and computational modeling of striatal-cortical function. Human Brain Mapping 30, 1963–1976.
Satoh, T., Nakai, S., Sato, T., Kimura, M., 2003. Correlated coding of motivation and outcome of decision by dopamine neurons. Journal of Neuroscience 23, 9913–9923.
Savastano, H.I., Miller, R.R., 1998. Time as content in pavlovian conditioning. Behavioural Processes 44, 147–162.
Schmajuk, N.A., 1997. Animal learning and cognition: a neural network approch. In: Problems in the Behavioural Sciences, Cambridge University Press.
Schneiderman, N., 1966. Interstimulus interval function of the nictitating membrane response of the rabbit under delay versus trace conditioning. Journal of Comparative and Physiological Psychology 62, 397–402.
Schoenbaum, G., Chiba, A.A., Gallagher, M., 1999. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. Journal of Neuroscience 19, 1876–1884.
Schoenbaum, G., Setlow, B., Saddoris, M.P., Gallagher, M., 2003. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron 39, 855–867.
Schultz, W., 1998. Predictive reward signal of dopamine neurons. Journal of Neurophysiology 80, 1.
Schultz, W., Apicella, P., Ljungberg, T., 1993a. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. Journal of Neuroscience 13, 900–913.
Schultz, W., Apicella, P., Scarnati, E., Ljungberg, T., 1993b. Neuronal activity in monkey ventral striatum related to the expectation of reward. Journal of Neuroscience 12, 4595–4610.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593.
Schultz, W., Dickinson, A., 2000. Neuronal coding of prediction errors. Annual Review of Neuroscience 23, 473–500.
Schultz, W., Romo, R., 1990. Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. Journal of Neurophysiology 63 (3), 607–624.
Semba, K., Fibiger, H., 1992. Afferent connections of the laterodorsal and the pedunculopontine tegmental nuclei in the rat: a retro- and antero-grade transport and immunohistochemical study. Journal of Comparative Neurology 323, 387–410.
Shepard, P.D., Holcomb, H.H., Gold, J.M., 2006. The presence of absence: habenular regulation of dopamine neurons and the encoding of negative outcomes. Schizophrenia Bulletin 32 (3), 417–421.
Smith, A.D., Bolam, J.P., 1990. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. Trends in Neuroscience 13, 259–265.
Smith, M.C., 1968. CS–US interval and us intensity in classical conditioning of the rabbit's nictitating membrane response. Journal of Comparative and Physiological Psychology 66 (3), 679–687.
Sokolov, E.N., 1963. Higher nervous functions: the orienting reflex. Annual Review of Physiology 25, 545–580.
Steinfels, G.F., Heym, J., Strecker, R.E., Jacobs, B.L., 1983. Response of dopamine neurons in cat to auditory stimuli presented across the sleep-waking cycle. Brain Research 277, 150–154.
Stuber, G., Klankder, M., de Ridder, B., Bowers, M.S., Joosten, R.N., Feenstra, M.G., Bonci, A., 2008. Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. Science 321, 1690–1692.
Suri, R.E., Schultz, W., 1999. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience 91, 871–890.
Suri, R.E., Schultz, W., 2001. Temporal difference model reproduces anticipatory neural activity. Neural Computation 13, 841.
Surmeier, D.J., Ding, J., Day, M., Wang, Z., Shen, W., 2007. D1 and d2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends in Neurosciences 30 (5), 228–235.
Sutton, R.S., 1988. Learning to predict by the method of temporal differences. Machine Learning 3, 9–44.
Sutton, R.S., Barto, A.G., 1990. Time-derivative models of pavlovian reinforcement. In: Moore, J.W., Gabriel, M. (Eds.), Learning and Computational Neuroscience. MIT Press, Cambridge, MA, pp. 497–537.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
Takayama, K., Miura, M., 1991. Glutamate-immunoreactive neurons of the central amygdaloid nucleus projecting to the subretrofacial nucleus of shr and wky rats: a double-labeling study. Neuroscience Letters 134, 62–66.
Tan, D., Bullock, C.O., 2008. A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. Journal of Neuroscience 28 (40), 10062–10074.
Terrace, H.S., 1981. Introduction: autoshaping and two-factor learning theory. In: Locurto, C.M., Terrace, H.S., Gibbon, J. (Eds.), Autoshaping and Conditioning Theory. Academic Press, New York, pp. 1–18.
Thorndike, E.L., 1911. Animal Intelligence: Experimental Studies. MacMillan Press.
Tobler, P.N., Dickinson, A., Schultz, W., 2003. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. Journal of Neuroscience 23, 10402–10410.
Tomie, A., 1981. Effect of unpredictable food on the subsequent acquisition of autoshaping: analysis of the context-blocking hypothesis. In: Locurto, C.M., Terrace, H.S., Gibbon, J. (Eds.), Autoshaping and Conditioning Theory. Academic Press, New York, pp. 181–215.
Ungless, M.A., Magill, P.J., Bolam, J.P., 2004. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. Science (New York, NY) 303, 2040–2042.
Waelti, P., Dickinson, A., Schultz, W., 2001. Dopamine responses comply with basic assumptions of formal learning theory. Nature 412, 43–48.
Wallace, D.M., Magnuson, D.J., Gray, T.S., 1992. Organization of amygdaloid projections to brainstem dopaminergic, noradrenergic, and adrenergic cell groups in the rat. Brain Research Bulletin 28, 447–454.
Wasserman, E.A., 1981. Response evocation in autoshaping: contributions of cognitive and comparative-evolutionary analyses to an understanding of directed action. In: Locurto, C.M., Terrace, H.S., Gibbon, J. (Eds.), Autoshaping and Conditioning Theory. Academic Press, New York, pp. 21–54.
Weible, A.P., McEchron, M.D., Disterhoft, J.F., 2000. Cortical involvement in acquisition and extinction of trace eyeblink conditioning. Behavioral Neuroscience 114, 1058–1067.
Wessells, M.G., 1974. The effects of reinforcement upon the prepecking behaviors in pigeons in the autoshaping experiment. Journal of the Experimental Analysis of Behavior 21, 125–144.
Wickens, J., Kotter, R., 1995. Cellular models of reinforcement. In: Houk, J.C., Davis, J.L., Beiser, D.G. (Eds.), Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, MA, pp. 187–214.
Wickens, J.R., Begg, A.J., Arbuthnott, G.W., 1996. Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. Neuroscience 70, 1–5.
Wickens, J.R., Horvitz, J.C., Costa, R.M., Killcross, S., 2007. Dopaminergic mechanisms in actions and habits. Journal of Neuroscience 27 (31), 8181–8183.
Wittmann, B.C., Daw, N.D., Seymour, B., Dolan, R.J., 2008. Striatal activity underlies novelty-based choice in humans. Neuron 58 (6), 967–973.