Expectancy, Ambiguity, and Behavioral Flexibility: Separable and Complementary Roles of the Orbital Frontal Cortex and Amygdala in Processing Reward Expectancies

Wolfgang M. Pauli, Thomas E. Hazy, and Randall C. O'Reilly

Abstract

■ Appetitive goal-directed behavior can be associated with a cue-triggered expectancy that it will lead to a particular reward, a process thought to depend on the OFC and basolateral amygdala complex. We developed a biologically informed neural network model of this system to investigate the separable and complementary roles of these areas as the main components of a flexible expectancy system. These areas of interest are part of a neural network with additional subcortical areas, including the central nucleus of amygdala, ventral (limbic) and dorsomedial (associative) striatum. Our simulations are consistent with the view that the amygdala maintains Pavlovian associations through incremental updating of synaptic strength and that

the OFC supports flexibility by maintaining an activation-based working memory of the recent reward history. Our model provides a mechanistic explanation for electrophysiological evidence that cue-related firing in OFC neurons is nonselectively early after a contingency change and why this nonselective firing is critical for promoting plasticity in the amygdala. This ambiguous activation results from the simultaneous maintenance of recent outcomes and obsolete Pavlovian contingencies in working memory. Furthermore, at the beginning of reversal, the OFC is critical for supporting responses that are no longer inappropriate. This result is inconsistent with an exclusive inhibitory account of OFC function. ■

INTRODUCTION

Deciding on which course of action to take critically depends on which rewards are expected to be available in the current situation. A reward may be expected because of the presence of a sensory stimulus that has reliably preceded reward delivery before or because reward was recently received in the current context. Existing data suggest that the amygdala learns which conditioned stimuli (CSs) and unconditioned stimuli (USs) are associated (Schoenbaum, Chiba, & Gallagher, 1999; Kita, Nishijo, Eifuku, Terasawa, & Ono, 1995) and that the orbital frontal cortex (OFC) can keep track of recent reward history in the current context (Wallis, 2007; Frank & Claus, 2006). We propose a model in which learning (and memory) in amygdala (specifically its basolateral complex, BLA) is solely weight-based, but memory in OFC is activation-based, and therefore does not depend on synaptic plasticity over relatively short time scales (O'Reilly & Munakata, 2000; see O'Reilly, Mozer, Munakata, & Miyake, 1999, for a theoretical discussion). Hence, OFC is capable of dynamically updating to new reward expectancy representations very quickly (despite no weight changes), but the amygdala changes more slowly because it is dependent on adapting its synaptic weights. As a result

of this dynamic, OFC supports flexible decision-making when and if environmental contingencies change.

To investigate the separable contributions of these different brain areas to reinforcement expectancies, we developed a biologically informed neural network model. This model captures empirical data on the effects of lesions of OFC, BLA, and simultaneous lesions of both areas on the ability to acquire the initial Pavlovian contingencies and the ability to adapt to a reversal of the Pavlovian contingencies.

Our simulations provide a mechanistic account for how OFC supports behavioral flexibility when Pavlovian contingencies change: Associations in BLA only change relatively slowly, but OFC can actively maintain a working memory of the recent reward history (Wallis, 2007; Frank & Claus, 2006). Thus, at the beginning of reversal, OFC can promote flexibility by biasing approach behavior associated with a recently experienced and now maintained US, even in the face of a CS that had previously predicted an aversive US. This role in Pavlovian reversal is in contrast with perspectives that have proposed that the primary role of OFC is the inhibition of inappropriate behavior (e.g., Elliott, Dolan, & Frith, 2000; Dias, Robbins, & Roberts, 1996; Damasio, 1994; Mishkin, 1964; Ferrier, 1886).

In addition to accounting for behavioral effects of lesions, the model provides an explicit mechanistic explanation for electrophysiological data that suggest that, when

University of Colorado at Boulder

Pavlovian contingencies change, OFC promotes behavioral flexibility by providing an ambiguous reinforcement expectancy. Under these circumstances, OFC shows nonselective cue-evoked activity (Schoenbaum, Roesch, Stalnaker, & Takahashi, 2009). The results of our simulations suggest that this nonselective activity could be because of the OFC simultaneously maintaining a working memory of the recent reward history, as well as the now-obsolete reinforcement expectancies being driven by the lagging BLA. Furthermore, consistent with empirical data, the speed at which the BLA can acquire and update Pavlovian associations is severely impaired if the OFC is lesioned data (Saddoris, Gallagher, & Schoenbaum, 2005). The model provides an explicit mechanistic explanation for this phenomenon as well, and it makes several related predictions.

METHODS

In our model, acquisition and performance of a Pavlovian approach/avoid task is the result of a division of labor among an expectancy system, an actor, and a critic. Different groups of layers in our model contribute to these three systems. The expectancy system produces a CS-evoked expectation for future reinforcement. It comprises the OFC, BLA, and ventral (limbic) striatum (VS). The actor system executes approach behavior toward appetitive reinforcers and avoidance behavior away from aversive reinforcers and comprises dorsomedial striatum (DMS) and motor cortices. The midbrain dopamine system and the central nucleus of amygdala (CNA) cause a phasic modulation of striatal dopamine and act as a critic that mediates feedback about the success of behavior (see Figure 1).

As a part of the expectancy system, the BLA has preexisting representations of USs and learns which CSs and USs are associated with each other (LeDoux, 2000; Schoenbaum et al., 1999; Kita et al., 1995; Quirk, Repa, & LeDoux, 1995). It is believed that this learning depends on synaptic plasticity (Fanselow & LeDoux, 1999) and is simulated by incremental updates of synaptic weights of projections from the CS to the BLA layer of the model. As in the Rescorla–Wagner model, learning in the BLA occurs when the US received at the end of a trial does not match the expectation developed at the time of the presentation of the CS (Rescorla & Wagner, 1972). Activation in BLA at CS onset is a combination of the result of this learning process and, to a limited extent, also the result of a top–down bias from the OFC (Corbit, Muir, & Balleine, 2003), which maintains a working memory of the recent reward history (Wallis, 2007; Frank & Claus, 2006).

The OFC of the model also has preexisting representations of USs (Ongür & Price, 2000). In contrast to the BLA, however, activation of US representations in the OFC is not a result of CS–US associations by neurons in this area (Holland & Gallagher, 2004). Instead this area is specialized for representing objects as USs and can also act as "working memory for USs," including which US was predicted by the BLA at CS onset (Wallis, 2007). It receives

this information over ascending projections from the BLA (Corbit et al., 2003). In addition, the OFC of the model can maintain a working memory of which USs have recently been received (Wallis, 2007; Frank & Claus, 2006). That is, if, for example, sucrose is presented to the model at the end of a trial, OFC can maintain this in working memory. Thus, active maintenance of USs can get into OFC working memory in two ways: (1) recent experience of a US, particularly if unexpected, and (2) expected USs based on BLA-driven inputs. Finally, there is no passive decay of working memory representations considered over the time scale of the tasks modeled nor an integration with previous reward history as in previous models on OFC function (e.g., Frank & Claus, 2006).

The OFC working memory mechanisms were developed using the PFC BG Working Memory framework (PBWM; Hazy, Frank, & O'Reilly, 2006, O'Reilly & Frank, 2006). The central tenet of the PBWM model is that the BG provides an adaptive, dynamic gating signal for controlling the active maintenance and updating, and the output of, information in frontal cortex (O'Reilly, 2006). The layers are interconnected with frontal cortex through a series of parallel loops (Postuma & Dagher, 2006; Middleton & Strick, 2000; Alexander, DeLong, & Strick, 1986). These loops enable the BG to exert a gating-like modulation of representations in frontal areas (see Figure 2). This kind of gating mechanism is consistent with a wide range of
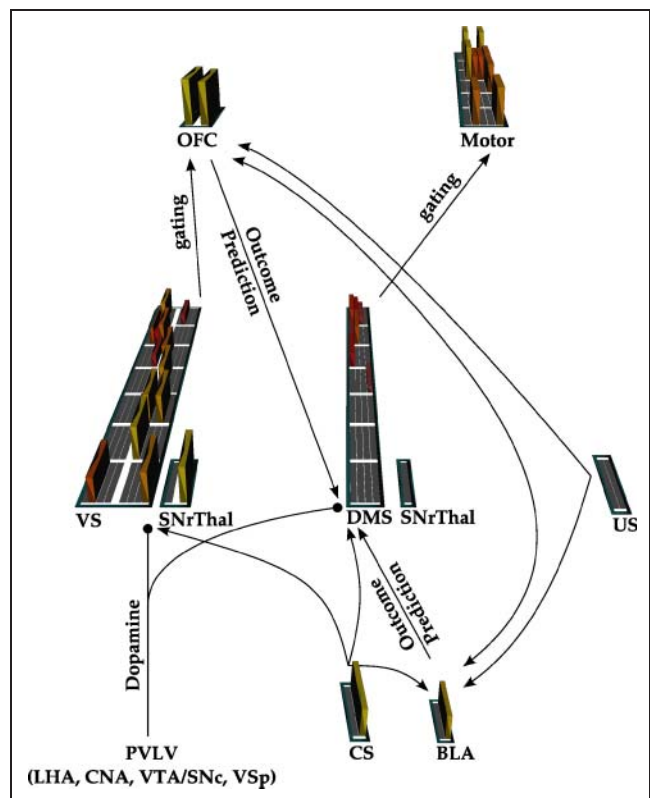

Figure 1. The model as implemented in the neural network simulator Emergent. LHA = lateral hypothalamus; VSp = patch-like neurons in VS.

Figure 2. The BGs are interconnected with frontal cortex. Working backward from the thalamus, which is bidirectionally excitatory with frontal cortex, the SNr is tonically active and inhibiting this excitatory circuit. When direct go pathway neurons in the striatum fire, they inhibit the SNr and thus disinhibit frontal cortex producing a gating-like modulation that we argue triggers the update of working memory representations in PFC. The indirect no-go pathway neurons of striatum counteract this effect by inhibiting the inhibitory GPe (globus pallidus, external segment; Hazy et al., 2007).
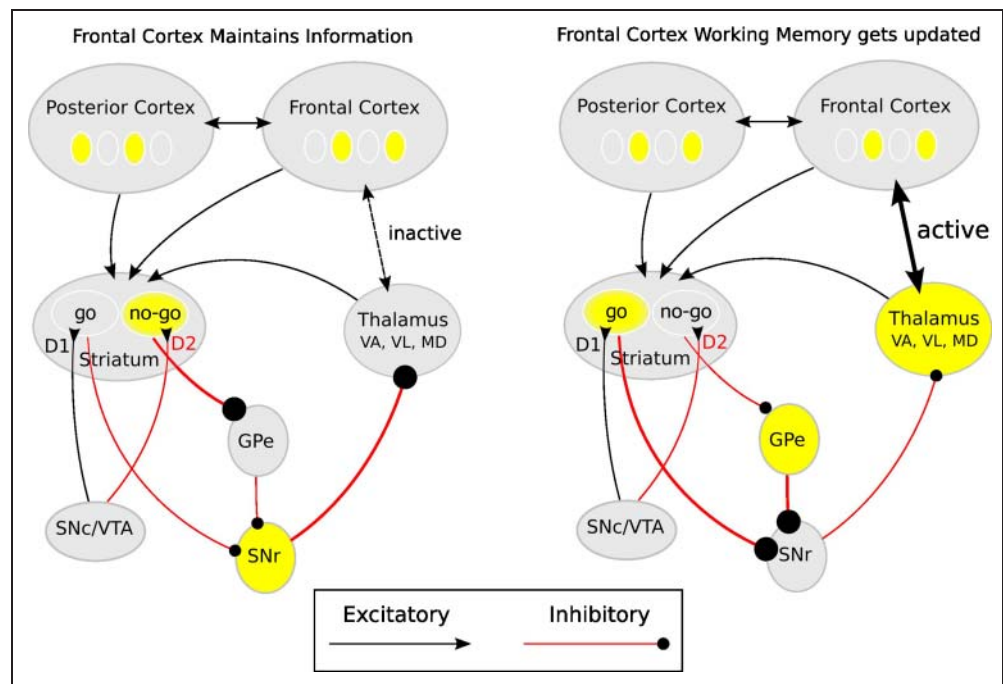
empirical data, and similar implementations of dynamic gating were included in previous computational models (e.g., Cisek, 2007; Houk et al., 2007; Humphries, Stewart, & Gurney, 2006; Frank, 2005; Brown, Bullock, & Grossberg, 2004; Gurney, Prescott, & Redgrave, 2001; Berns & Sejnowski, 1998; Mink, 1996; Dominey, Arbib, & Joseph, 1995; Houk, Adams, & Barto, 1995; Houk & Wise, 1995; Wickens, Kotter, & Alexander, 1995).

In our model, the VS layer provides a dynamic gating mechanism for the OFC (Frank & Claus, 2006). The VS learns to update memory in OFC based on excitatory input from the BLA and input from the CS layer (Cardinal, Parkinson, Hall, & Everitt, 2002; Gray, 1999). Learning when to update OFC depends on phasic dopamine release in VS by neurons in ventral tegmental area (VTA)/substantia nigra pars compacta (SNc; see below). Although OFC and BLA both receive sensory information from higher-level sensory areas of the temporal lobe (Ghashghaei & Barbas, 2002), the OFC of the model does not receive direct sensory input from the CS layer. Our model does not include these projections, because we believe that the extremely slow learning rate at OFC synapses precludes the formation of CS–US associations within the timescale of the experiments we simulated (Holland & Gallagher, 2004). That is, if the expectation for a particular reward is activated in OFC when a CS is presented, this is because BLA is sending its US prediction to OFC, but not because OFC itself learned the CS–US association (Schoenbaum, Setlow, Saddoris, & Gallagher, 2003).

With sufficient training, we do believe that OFC can eventually acquire multisensory feature–US associations, so that a US representation in OFC will include information about all the features reliably associated with the core sensory experience (Holland & Gallagher, 2004), which

may be the core process underlying stimulus substitution. Thus, the OFC can come to represent all of the multisensory aspects associated with a reward, such as its size, shape, texture, and flavor (Rolls & Grabenhorst, 2008; Schoenbaum & Roesch, 2005), and in our view, when it pairs sensory features with USs, it does so as unitary US representations and not as CS–US pairings per se. We do not address these aspects of OFC function in our simulations.

In our simulations of the function of the OFC, we focus on a lateral region for which there are strong anatomical and functional parallels between rodents and primates (see also Schoenbaum et al., 2009). This region encompasses lateral orbital regions, anterior parts of the agranular insular cortex and the dorsal bank of the rhinal sulcus in rodents. These areas are heavily interconnected with the BLA, VS, mediodorsal thalamus, and sensory cortices. These areas in rodent OFC correspond to Areas 11, 12, and 13 in the primate OFC (Schoenbaum & Roesch, 2005; Ongür, Ferry, & Price, 2003; Ongür & Price, 2000; Preuss, 1995). The role of this area in decision-making is to determine which stimulus outcomes are possible in the current context, but not what is necessary to achieve that outcome. More medial aspects of ventral frontal cortex are involved in learning and representing action–outcome values, including costs (Rushworth, Behrens, Rudebeck, & Walton, 2007). The division of labor among regions of OFC has been discussed elsewhere (Noonan et al., 2010; Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Ongür et al., 2003; Ongür & Price, 2000).

The actor system of the model consists a simulated DMS and motor cortices. It is well accepted that the DMS is involved in the initiation of the motor gating of behavior in motor cortices (Mink, 1996; Wickens, 1993). The DMS is believed to guide goal-directed behavior according to the

expectancy information it receives from the BLA and OFC (Pauli, Atallah, & O'Reilly, 2010; Pauli, Hazy, & O'Reilly, 2009; Balleine, Delgado, & Hikosaka, 2007). Lesions of this region have been found to lead to similar reversal deficits as lesions of the OFC itself (Clarke, Robbins, & Roberts, 2008). In our model, OFC and BLA can independently promote a go response (e.g., approach the food well) if they predict an appetitive US (Frank, Seeberger, & O'Reilly, 2004). If OFC and BLA expect an aversive US, they bias the DMS medium spiny neurons of the indirect (no-go) pathway to prevent approaching the food well. The DMS also receives sensory input from the CS layer (McGeorge & Faull, 1989). Because of this connection, the DMS can acquire CS–response associations (Everitt & Robbins, 2005), so that the conditioned response is spared even if both BLA and OFC are lesioned (Stalnaker, Franz, Singh, & Schoenbaum, 2007), and there is most likely no acquisition of Pavlovian CS–US associations.

For the above gating mechanism to work successfully, the striatum has to learn when to update representations in frontal areas. This learning is dopamine-based and allows each striatal projection neuron (medium spiny neuron) to develop its own unique pattern of input weights that determine its actions. Dopamine release in the striatum of our model is determined by projections from the dopaminergic neurons of the SNc/VTA, captured by the PVLV model (primary value, learned value; Hazy, Frank, & O'Reilly, 2010; Hazy et al., 2007; O'Reilly, Frank, Hazy, & Watz, 2007; O'Reilly & Frank, 2006).

It is well established that the midbrain dopamine neurons in the SNc/VTA of the mammalian brain are driven by inputs from the CNA, the lateral hypothalamus, and the patch-like neurons of the VS (Ahn & Phillips, 2003; Floresco, West, Ash, Moore, & Grace, 2003; Fudge & Haber, 2000; Joel & Weiner, 2000; Rouillard & Freeman, 1995; Semba & Fibiger, 1992). The contributions of these inputs are described by the PVLV model as follows (Hazy et al., 2007, 2010; O'Reilly et al., 2007; O'Reilly & Frank, 2006). The lateral hypothalamus delivers primary reward information and contributes to the phasic dopamine release in response to unexpected reward delivery. The patch-like neurons in the VS learn to expect such rewards and thereby block the dopamine spike that would otherwise occur to them. This is the PV system of PVLV. The LV system, involving the CNA, is important for learning reward associations for CSs, which can then drive dopamine firing at the time of CS onset. These two interacting systems provide a good account of the extant neural recording data from the SNc (Schultz, 1998; Schultz, Apicella, & Ljungberg, 1993). In many learning paradigms, the PVLV algorithm can be considered as a biologically informed version of the temporal differences algorithm (Sutton & Barto, 1998; Sutton, 1988), although there are also important differences between these two models in some specific learning paradigms (Hazy et al., 2010).

The functional contribution of the PVLV system is to provide positive dopamine bursts for successful behavior and CSs associated therewith and negative dopamine dips for unsuccessful behavior and associated CSs. The positive dopamine bursts cause go pathway neurons in the striatum to become more active (because of a preponderance of dopamine D1 receptors, which are excitatory) and no-go pathway neurons to become less active (from D2 receptors, which are inhibitory; Shen, Flajolet, Greengard, & Surmeier, 2008; Frank, 2005; Frank et al., 2004). The opposite case holds for negative dopamine dips. This shapes the gating firing in ways that lead to successful learning of complex working memory tasks in the PBWM model (Hazy et al., 2006, 2007; O'Reilly & Frank, 2006).

Because the main focus of our model was on the acquisition and reversal of Pavlovian contingencies, we only simulated the effect of phasic dopamine on plasticity in striatal areas but did not simulate modulations of tonic dopamine levels in the BLA and frontal cortex. Modulation of tonic dopamine levels in BLA is thought to be critical for motivation tone (Niv, Daw, Joel, & Dayan, 2007) and the generalized form of Pavlovian-to-instrumental transfer (Hazy et al., 2010). In PFC, dopamine has been proposed to affect the amount of information held in working memory buffers in PFC networks (Seamans & Yang, 2004).

Separable Functional Roles of BLA and CNA

The amygdala has long been recognized for its critical role in emotional processing (e.g., LeDoux, 2000; Adolphs, Tranel, Damasio, & Damasio, 1995; Quirk et al., 1995). Despite the dominant interest in the role of amygdala in fear and anxiety (Fanselow & Gale, 2003; Fanselow & LeDoux, 1999; Davis, 1992), its role in representing positive affect has started to receive more attention as well (Murray, 2007; Paton, Belova, Morrison, & Salzman, 2006; Gottfried, O'Doherty, & Dolan, 2003). The CNA and BLA of the amygdala have been shown to be highly dissociable across many experimental paradigms and, in our model, make separable contributions to goal-directed behavior. As a key component of the PVLV reinforcement learning system, the CNA (LV_e in PVLV) learns to control the release of dopamine in striatal layers at the onset of the CS (Hazy et al., 2007, 2010; O'Reilly et al., 2007; O'Reilly & Frank, 2006). The BLA, on the other hand, does not have direct access to the dopamine cells but does project very densely to the VS and associative striatum (i.e., DMS), which CNA does not. Thus, the BLA is in a position to influence the learning and performance of goal-directed behaviors by signaling its expectancies about particular USs to downstream areas (Hatfield, Han, Conley, & Holland, 1996).

Trial Structure

Each experimental trial corresponds to three discrete steps in our simulations. Each simulation step consists of one minus phase and one plus phase (O'Reilly, 1996b).

The first step is for "stimulus sampling," the CS is presented to the network until settling finishes. During this step, working memory in OFC can be updated. In the second "response" step, the model decides whether to approach or avoid the food well and receives simulated dopamine feedback for its choice. In the third "feedback" step, USs are presented according to which response the model chose in the "response" step. Working memory in OFC is updated to maintain the history for recent rewards. The distinction between "response" and "feedback" trials in the model is required to accommodate computational constraints associated with the PBWM mechanisms.

Training Parameters

The model first had to learn to associate one conditioned (CS1+) stimulus with an appetitive US and another CS (CS2−) with an aversive US. The model was trained until it had correctly performed 95 trials of each type without any errors. After the model had acquired the initial associations, contingencies were reversed so that the first CS was now associated with a negative US (CS1−) and the other with a positive US (CS2+). The model was trained on the reversed contingencies until it had performed 100 trials of each type without an error. The model was run with either set of lesions (BLA only, BLA+OFC lesion, and no lesion) for 50 runs to acquire a good estimate of the average performance.

Parameter Fitting

No systematic attempt was made to fit the exact quantitative pattern of the rat behavioral data. To capture the effects of lesions on acquisition on reversal performance, we adjusted the following parameters:

- The weight scale between the US and BLA layer so that BLA would learn more slowly which CSs and USs are associated if the OFC was lesioned. The value was reduced such that an US representation in the BLA was less active at the time of the US presentation without the additional excitatory input from the OFC.
- We increased the amount of Hebbian learning to make sure that the CS–US associations in BLA would strengthen further even when there was no error in the US expectation.
- We increased the learning rate of the CS layer to DMS projections so that the model would acquire the initial contingencies at the same rate if it was intact or if BLA and OFC were lesioned simultaneously.
- We increased the weight scale between the BLA, DMS, and OFC to DMS, so that expectations about USs of these two areas were able to exert a strong bias onto the DMS and overcome net input from the CS layer.
- We increased the random go firing in the DMS so that the model would start exploring faster at the

beginning of reversal after not receiving reward in either trial type for several trials in a row.

The emergent project file can be downloaded at http://grey.colorado.edu/CompCogNeuro/index.php/whip_ofc. Further details of the equations used, based on the Leabra unified framework for neural modeling (O'Reilly & Munakata, 2000), can be found in the Appendix.

RESULTS

We developed a biologically informed neural network model to investigate the role of the OFC and BLA in Pavlovian acquisition and reversal. To test the contributions of the two areas to the acquisition and reversal of Pavlovian contingencies, we trained the model to associate two CSs with two different USs. The model first had to learn to associate one CS with a positive US and another CS with a negative US. After the model had acquired the initial associations, contingencies were reversed so that the first CS was now associated with a negative US and the other with a positive US.

Reversal Deficit after OFC Lesions

OFC lesions have been repeatedly found to cause learning impairments if contingencies are reversed after acquisition in Pavlovian conditioning studies. As displayed in Figure 3, the model also exhibited a reversal deficit after inactivation of the OFC. Reversal deficits seem to be caused by perseverative encoding of the original Pavlovian CS–US associations in the BLA, which would normally be compensated for by activation-based working memory for recent outcomes in the OFC, as described earlier. Stalnaker et al. (2007) were able to confirm this idea by abolishing
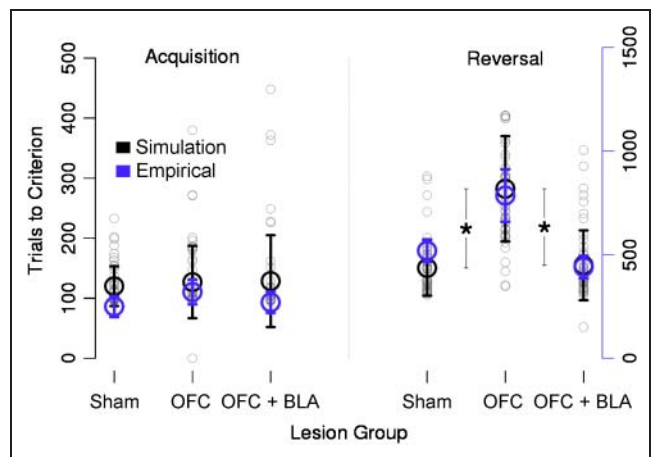

Figure 3. Trials to criterion until acquisition and reversal for the different lesion groups for the model. Paralleling empirical data (Stalnaker et al., 2007, Figure 2) neither of the lesion groups showed a deficit for the acquisition of the initial associations, and simultaneous lesions of OFC and BLA abolished the reversal deficit found after OFC lesions alone.

Figure 4. Selectivity of neuronal firing of neurons in OFC and BLA as a function of epoch since the beginning of acquisition and reversal, respectively. This value is positive if neurons of a particular US are more active if the CS that is associated with this US is presented (see Appendix for details). Line represents mean selectivity over the 50 runs of the model; area represents standard error. Progress refers to percent completion of either acquisition or reversal phase. SSE refers to sum of squared errors.
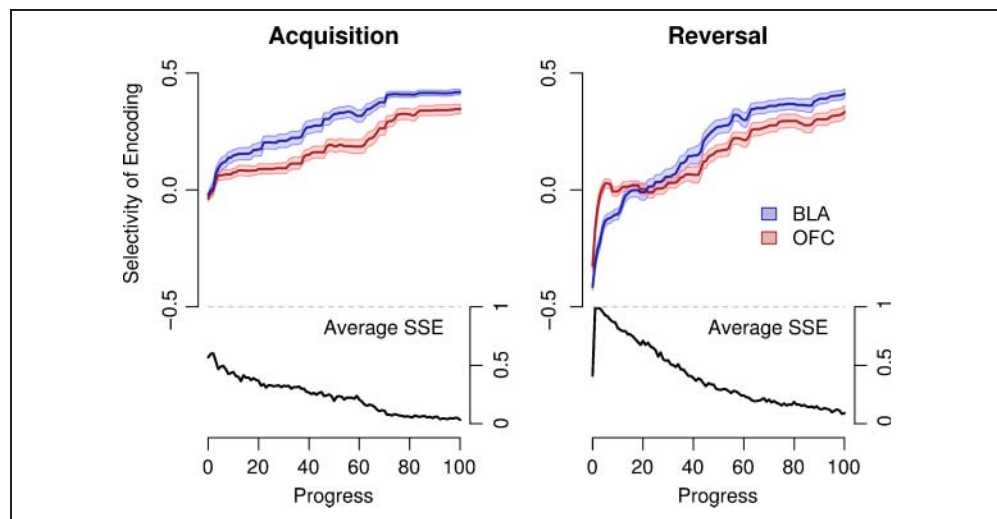
the reversal deficit by simultaneously ablating the OFC and the BLA in rats. Simultaneous lesions of BLA and OFC in our model also abolished the reversal deficits found after OFC lesions (Figure 3). In the case of simultaneous inactivation of OFC and BLA, phasic dopamine release in response to unexpected delivery of the positive US and phasic reductions of dopamine in response to the delivery of a negative US support the acquisition of CS–response associations in the DMS. With simultaneous OFC and BLA lesions, the model produces approach and avoid behavior without the expectancy for a particular US.

Ambiguous CS-evoked Activity in OFC

How does the OFC support behavioral flexibility? Rolls (1996) originally suggested that the OFC was fast and flexible at encoding CS–US associations and was therefore particularly critical when Pavlovian contingencies changed. Although Rolls (1996) originally attributed this flexibility to rapid weight-based learning, we have reframed this flexibility in terms of activation-based memory, as described earlier. According to either general framework, the OFC provides this updated associative information to other brain areas to guide appropriate behavior. Several single-unit studies provided evidence in support of this hypothesis (Schoenbaum et al., 1999; Rolls, Critchley, & Treves, 1997; Thorpe, Rolls, & Madison, 1983).

However, although the OFC learns to fire selectively in anticipation of a particular US (Schoenbaum, Chiba, & Gallagher, 1998), selective firing of OFC neurons neither develops particularly rapidly in comparison with other brain areas nor is it very pervasive (Paton et al., 2006; Stalnaker, Roesch, Franz, Burke, & Schoenbaum, 2006; Schoenbaum et al., 1999). The OFC and BLA layers in our model exhibit this same behavior. As shown in Figure 4, selective cue-related firing occurred earlier during acquisition and reversal in the BLA than in the OFC. Activation in the OFC layer represents a combination of both the current expectation by the BLA, but also recent reward history.

That is, as long as performance is low and unexpected USs will periodically be received, the OFC will maintain both received US as well as the (now incorrect) expected US in working memory, signaling that both of these USs are possible in the current context. In contrast to the slow development of selective anticipatory OFC activity, OFC already fires selectively at the moment a US is received early during acquisition and reversal (Figure 5).

OFC Modulates Plasticity in BLA

Acquisition of Pavlovian associations by the BLA has been found to depend on a functioning OFC (Saddoris et al., 2005). In particular, the lateral OFC seems to be more critical for learning than for decision-making directly (Noonan et al., 2010). If we lesioned the OFC in the model, anticipatory firing would only develop slowly, if at all, in the BLA of
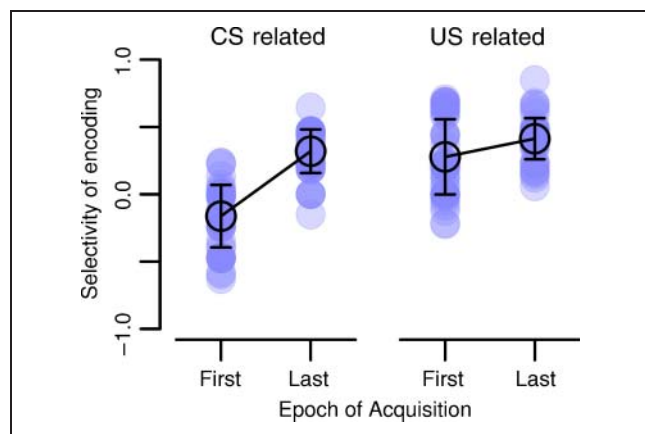

Figure 5. Selectivity of neuronal firing of OFC neurons at the beginning and end of acquisition of the initial Pavlovian contingencies. Even early in acquisition, OFC neurons fire selectively in response to the delivery of a US. This is the result of preexisting orbito-frontal representations of USs. In contrast, neurons do not fire in expectation of a particular US when a CS is presented early in acquisition but acquires this selectivity toward the end of the acquisition.
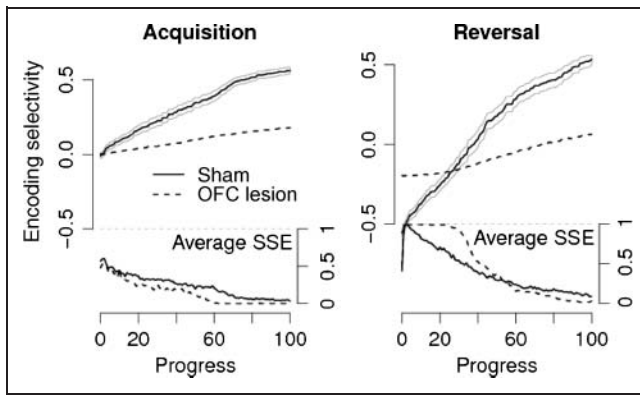
Figure 6. OFC lesions impair acquisition and reversal of Pavlovian associations by the BLA. Graphs show selectivity of cue-evoked activity in BLA with lesion to the OFC (solid) and no lesion to the OFC (dashed). Areas indicate standard error; SSE refers to sum of squared errors.

the model because of a reduced excitatory input to BLA neurons representing the delivered US (Figure 6). On the other hand, the Pavlovian associations in BLA developed more strongly if OFC activity at CS onset was ambiguous or incorrect (Figure 7). If OFC represents an incorrect US outcome expectation or that both US outcomes are possible, when a CS is presented, the top–down bias from OFC onto the BLA will also cause BLA to have this incorrect expectation. When the actual US is presented at the end of the trial, the difference between the US expectation and the actual US will increase the amount of plasticity in BLA. That is, the expectancy error in the OFC activation is not only proportional to the amount of learning in BLA, it actually causes a modulation of plasticity in BLA. This is consistent with the finding that animals are better at adapting to a reversal of Pavlovian contingencies if selective firing in response to a CS in OFC is slow at reversing (Stalnaker et al., 2006).

This modulation of plasticity by the expectancy error is similar to the finding that the phasic changes in dopamine release proportional to reward prediction error modulate plasticity in striatal areas. However, unlike the activity in midbrain dopamine neurons, activity in OFC in response to the delivery of a US is not modulated by how much this US had been expected (Schoenbaum et al., 2003; Takahashi et al., 2009, Figure 5).

DISCUSSION

We were able to develop a computational model that captures various findings of studies that looked at electrophysiological changes in the BLA and the OFC and the effects of lesions to either area on the ability to acquire and adapt to the reversal of Pavlovian contingencies. These simulations were based on two simple assumptions. The first assumption was that the BLA learns CS–US associations via purely weight-based learning and, thus, predict USs on the basis of CS cues; the second was that the OFC

acts as "working memory for USs" based on activation-based memory—and US representations can get loaded into OFC in two ways: (1) when USs occur and (2) when BLA predicts them.

We were able to account for the finding that neither lesions of the OFC or the BLA nor simultaneous lesions of both areas would impair initial performance in this catergory of tasks (Stalnaker et al., 2007). Furthermore, our simulations captured the empirical finding that lesions of the OFC alone would greatly impair reversal performance whereas simultaneous lesions of both areas would abolish this reversal deficit (Stalnaker et al., 2007). Although neither lesion affected behavior during the acquisition phase, the BLA only acquired the initial Pavlovian contingencies very slowly, if the OFC was lesioned in the model. At the beginning of the reversal phase, the BLA then contributed a bias to behavior according to the initial contingencies, which were no longer appropriate. The OFC supported rapid reversal in two different ways. First, it supported rapid reversal of associative weights in the BLA. Secondly, it maintained recent trial USs in working memory and, therefore, biased responding in DMS against the no-longer-appropriate Pavlovian associations stored in the BLA.

The model could solve Pavlovian acquisition and reversal without any contributions from the expectancy system when OFC and BLA were lesioned simultaneously, which is consistent with empirical data (Stalnaker et al., 2007). With simultaneous lesions, the model solved the task because the DMS would acquire stimulus–response associations through reinforcement learning. In a sense, the model was producing the correct behavior without anticipating a particular US to result from it, that is, without acquiring Pavlovian CS–US associations. Although the expectancy system would normally be involved in this task, this demonstrates the task can be solved without the acquisition of CS–US associations. That the simultaneous lesions did not affect the speed of acquisition or reversal appears to be rather perplexing, because it implies that
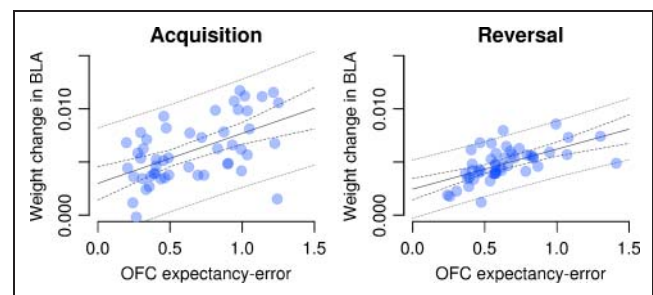
Figure 7. The speed of learning BLA is directly proportional to the error in the US expectation of the OFC during acquisition (left) and reversal of Pavlovian contingencies (right). The OFC expectancy error refers to the sum squared error between OFC activation at the time of the CS and the time of the US. Solid line represents prediction of the linear model, dashed line represents prediction interval, and dotted line represents confidence interval.

the expectancy system is not really very useful. However, we interpret this to be because of the extremely impoverished environment, in which there are only two things to do (avoid and approach). This is similar to findings that simple instrumental tasks (e.g., *t* maze) can be learned without the DMS, because the task is so simple that the dorsal striatum can just acquire S–R associations (Palencia & Ragozzino, 2005; Featherstone & McDonald, 2004). We believe that the expectancy system becomes more critical when there are multiple options to choose from within the same valence category (e.g., R1–sugar, R2–food pellet). Animals cannot learn those tasks without, for example, the DMS, as the actor of the expectancy system (Yin, Ostlund, Knowlton, & Balleine, 2005). More generally, the expectancy system is critical for modulating behavior as a function of changing needs and goals, as explored, for example, in devaluation and related paradigms (Ostlund & Balleine, 2007).

In addition to capturing these behavioral findings, our model also accounted for various electrophysiological findings. Consistent with empirical data, BLA acquired and reversed Pavlovian associations more slowly, if the OFC layer was lesioned in the model (Saddoris et al., 2005). Furthermore, if OFC was slow to adapt to the contingency reversal or failed to do so altogether, the BLA would acquire the reversed Pavlovian associations more readily (Stalnaker et al., 2006).

It has previously been proposed that the OFC inhibits inappropriate responses (Elliott et al., 2000; Dias et al., 1996; Damasio, 1994; Mishkin, 1964; Ferrier, 1886). This inhibitory role of the OFC is consistent with deficits in detour reaching tasks (Wallis, Dias, Robbins, & Roberts, 2001) and stop signal tasks (Eagle et al., 2008). However, other studies have also produced results inconsistent with an inhibitory role of OFC. For example, rhesus monkeys with orbito-frontal lesions were still capable of inhibiting a prepotent response to pick up a small reward to receive a larger reward later (Chudasama, Kralik, & Murray, 2007). The contribution of the OFC in our model is also inconsistent with an exclusive role of OFC in response inhibition. At the beginning of reversal, the model continues to approach in response to the presentation of one CS (CS1), because it learned to associate it with the positive US during acquisition. Every time it approaches the food well in response to the CS1, the model receives the aversive US and finally stops responding to CS1. Because it had also previously learned that CS2 is associated with an aversive US, it never approaches the food well in response to CS2 and, in fact, stops behaving completely. Thus, it never gets an opportunity to experience the new contingencies—until the model eventually starts to explore again after not receiving any positive US for several trails. As soon as this happens, OFC holds on to the positive US and exerts a bias onto the DMS that makes this approach behavior more likely to be expressed again. Taken together, therefore, we believe that converging evidence supports the "working memory for USs" model

of OFC function. Although our simulations focused on the role of lateral OFC in Pavlovian, we believe that they provide a more general and comprehensive description of the OFC's role in supporting flexible behavior that goes beyond inhibition of inappropriate behaviors (Schoenbaum et al., 2009).

Predictions from the Model

The model makes several testable predictions:

- Stalnaker et al. (2007) found that simultaneous lesions of the BLA and the OFC would abolish the reversal deficits associated with OFC lesions alone. If both layers were inactivated in our model, it would solve the task according to stimulus–response associations in the DMS. We predict that, if plasticity is blocked in the DMS throughout the reversal period, animals with simultaneous lesions to OFC and BLA should be significantly impaired because the DMS would be unable to reverse the stimulus-response associations.

- The second prediction is based on the fact that the model was able to account for the above-discussed empirical findings without requiring any synaptic plasticity in OFC. Thus, blocking plasticity in OFC at any point during the experiments without blocking neuronal activity, for example, by injecting the selective PKM_{zeta} inhibitor ZIP (see Sacktor, 2011), should not affect the results, in particular, the speed of reversal learning.

- If we lesioned the OFC of the model, the BLA would very slowly acquire the initial Pavlovian contingencies and provide an inappropriate response bias onto the DMS at the beginning of the reversal period that would impair the ability to adapt to the changed contingencies. Because lesions to the BLA did not affect the ability of rats and the model to acquire the initial Pavlovian contingencies (because of CNA being intact), blocking plasticity in the BLA should not affect the ability of animals to acquire Pavlovian contingencies either (as already shown for BLA lesions) and may actually facilitate the adaptation to a reversal of Pavlovian contingencies because the BLA would not contribute the now-inappropriate Pavlovian contingencies.

- Finally, blocking plasticity in BLA should prevent selective firing in the OFC at the CS onset, because the OFC would then be lacking the CS–US associations normally acquired by the BLA and the OFC does not learn fast enough on its own.

Conclusions

Our simulations suggest a division of labor within an expectancy system between the OFC and the BLA. The BLA acquires Pavlovian associations based on long-term synaptic

plasticity. The OFC supports flexibility by maintaining activation-based memories for USs, including the recent reward history. This memory does not require synaptic plasticity. Therefore, the OFC is a source of flexibility, and the BLA is a source of continuity. Ambiguous reward expectancies in OFC at the time of the CS presentation promotes behavioral flexibility and synaptic plasticity in the BLA. When contingencies change, OFC supports responses that are no longer inappropriate, which is inconsistent with an exclusive inhibitory role of OFC function.

APPENDIX: IMPLEMENTATIONAL DETAILS

The model was implemented using the Leabra framework, which is described in detail in O'Reilly (2001) and O'Reilly and Munakata (2000) and summarized here. See Table 1 for a listing of parameter values; nearly all of which are at their default settings. These same parameters and equations have been used to simulate over 40 different models in O'Reilly and Munakata (2000) and a number of other research models. Thus, the model can be viewed

Table 1. Parameters for the Simulation

Parameter	Value
E_l	0.15
E_i	0.15
E_e	1.00
V_{rest}	0.15
τ	.02
k In/Out	1
k PFC	4
k PVLV	8
K_{hebb}	.01
to PFC k_{hebb}	.001*
$\overline{g_l}$	0.10
$\overline{g_i}$	1.0
$\overline{g_e}$	1.0
Θ	0.25
γ	600
k Hidden	7
k Striatum	1
ϵ	.01
to PFC ϵ	.001*

See equations in text for explanations of parameters. All are standard default parameter values except for those with an asterisk. The slower learning rate of PFC connections produced better results and is consistent with a variety of converging evidence, suggesting that PFC learns more slowly than the rest of cortex (Morton & Munakata, 2002).

as an instantiation of a systematic modeling framework using standardized mechanisms instead of constructing new mechanisms for each model. The model can be obtained by emailing the first author at oreilly@psych.colorado.edu.

Pseudocode

The pseudocode for Leabra is given here, showing exactly how the pieces of the algorithm described in more detail in the subsequent sections fit together.

Outer loop: Iterate over events (trials) within an epoch. For each event:

1. Iterate over minus (−), plus (+), and update (++) phases of settling for each event.

   (a) At start of settling:

      i. For non-PFC/BG units, initialize state variables (activation, v_m, etc.).
      ii. Apply external patterns (clamp input in minus, input and output, external reward based on minus-phase outputs).

   (b) During each cycle of settling, for all nonclamped units:

      i. Compute excitatory netinput ($g_e(t)$ or η_j, Equation 2; Equation 21 for SNr/Thal units).
      ii. For striatum go/no-go units in ++ phase, compute additional excitatory and inhibitory currents based on dopamine inputs from SNc (Equation 20).
      iii. Compute kWTA inhibition for each layer, based on g_i^{Θ} (Equation 6):

         A. Sort units into two groups based on g_i^{Θ}: top k and remaining $k + 1$ to n.
         B. If basic, find k and $k + 1$-th highest; if average based, compute average of $1 \rightarrow k$ and $k + 1 \rightarrow n$.
         C. Set inhibitory conductance g_i from g_k^{Θ} and g_{k+1}^{Θ} (Equation 5).

      iv. Compute point neuron activation combining excitatory input and inhibition (Equation 1).

   (c) After settling, for all units:

      i. Record final settling activations by phase (y_j^-, y_j^+, y^{++}).
      ii. At end of + and ++ phases, toggle PFC maintenance currents for stripes with SNr/Thal act > threshold (.1).

2. After these phases, update the weights (based on linear current weight values):

   (a) For all non-BG connections, compute error-driven weight changes (Equation 8) with soft weight bounding (Equation 9), Hebbian weight changes

from plus-phase activations (Equation 7), and overall net weight change as weighted sum of error-driven and Hebbian (Equation 10).

(b) For PV units, weight changes are given by delta rule computed as difference between plus phase external reward value and minus phase expected rewards (Equation 11).

(c) For LV units, only change weights (using Equation 13) if PV expectation $> \theta_{pv}$ or external reward/punishment actually delivered.

(d) For striatum units, weight change is the delta rule on dopamine-modulated second-plus phase activations minus unmodulated plus phase acts (Equation 19).

(e) Increment the weights according to net weight change.

Point Neuron Activation Function

Leabra uses a point neuron activation function that models the electrophysiological properties of real neurons while simplifying their geometry to a single point. The membrane potential V_m is updated as a function of ionic conductances g with reversal (driving) potentials E as follows,

$$\Delta V_m(t) = \tau \sum_c g_c(t)\overline{g_c}(E_c - V_m(t)) \qquad (1)$$

with three channels (c) corresponding to e as excitatory input, l as leak current, and I as inhibitory input. Following electrophysiological convention, the overall conductance is decomposed into a time-varying component $g_c(t)$ computed as a function of the dynamic state of the network and a constant $\overline{g_c}$ that controls the relative influence of the different conductances.

The excitatory net input/conductance $g_e(t)$ or η_j is computed as the proportion of open excitatory channels as a function of sending activations times the weight values,

$$\eta_j = g_e(t) = \langle x_i w_{ij} \rangle = \frac{1}{n}\sum_i x_i w_{ij} \qquad (2)$$

The inhibitory conductance is computed via the k-winners-take-all (kWTA) function described in the next section, and leak is a constant.

Activation communicated to other cells (y_j) is a thresholded (Θ) Sigmoidal function of the membrane potential with gain parameter γ:

$$y_j(t) = \frac{1}{\left(1 + \frac{1}{\gamma[V_m(t) - \Theta]_+}\right)} \qquad (3)$$

where $[x]_+$ is a threshold function that returns 0 if $x < 0$ and x if $X > 0$. Note that if it returns 0, we assume $y_j(t) = 0$, to avoid dividing by 0. To produce a less discontinuous deterministic function with a softer threshold, the function is convolved with a Gaussian noise kernel ($\mu = 0$, $\sigma = .005$), which reflects the intrinsic processing noise of biological neurons,

$$y_j^*(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-z^2/(2\sigma^2)}y_j(z - x)\mathrm{d}z \qquad (4)$$

where x represents the $[V_m(t) - \Theta]_+$ value and $y_j^*(x)$ is the noise-convolved activation for that value. In the simulation, this function is implemented using a numerical lookup table.

kWTA

Leabra uses a kWTA function to achieve inhibitory competition among units within a layer (area). The kWTA function computes a uniform level of inhibitory current g_i for all units in the layer, such that the $k + 1$-th most excited unit within a layer is generally below its firing threshold whereas the kth is typically above threshold,

$$g_i = g_{k+1}^{\Theta} + q(g_k^{\Theta} - g_{k+1}^{\Theta}) \qquad (5)$$

where $0 < q < 1$ (.25 default used here) is a parameter for setting the inhibition between the upper bound of g_k^{Θ} and the lower bound of g_{k+1}^{Θ}. These boundary inhibition values are computed as a function of the level of inhibition necessary to keep a unit right at threshold,

$$g_i^{\Theta} = \frac{g_e^* \overline{g}_e(E_e - \Theta) + g_l \overline{g}_l(E_l - \Theta)}{\Theta - E_i} \qquad (6)$$

where g_e^* is the excitatory net input without the bias weight contribution—this allows the bias weights to override the kWTA constraint.

In the basic version of the kWTA function, which is relatively rigid about the kWTA constraint and is therefore used for output layers, g_k^{Θ} and g_{k+1}^{Θ} are set to the threshold inhibition value for the kth and $k + 1$-th most excited units, respectively. In the average-based kWTA version, g_k^{Θ} is the average g_i^{Θ} value for the top k most excited units, and g_{k+1}^{Θ} is the average of g_i^{Θ} for the remaining $n - k$ units. This version allows for more flexibility in the actual number of active units, depending on the nature of the activation distribution in the layer.

Hebbian and Error-driven Learning

For learning, Leabra uses a combination of error-driven and Hebbian learning. The error-driven component is the symmetric midpoint version of the GeneRec algorithm

(O'Reilly, 1996a), which is functionally equivalent to the deterministic Boltzmann machine and contrastive Hebbian learning. The network settles in two phases, an expectation (minus) phase where the network's actual output is produced and an outcome (plus) phase where the target output is experienced, and then computes a simple difference of a pre- and postsynaptic activation product across these two phases. For Hebbian learning, Leabra uses essentially the same learning rule used in competitive learning or mixtures of Gaussians, which can be seen as a variant of the Oja normalization (Oja, 1983). The error-driven and Hebbian learning components are combined additively at each connection to produce a net weight change.

The equation for the Hebbian weight change is

$$\Delta_{\text{hebb}} w_{ij} = x_i^+ y_j^+ - y_j^+ w_{ij} = y_j^+ \left(x_i^+ - w_{ij} \right) \qquad (7)$$

and for error-driven learning using contrastive Hebbian learning,

$$\Delta_{\text{err}} w_{ij} = \left(x_i^+ y_j^+ \right) - \left(x_i^- y_j^- \right) \qquad (8)$$

which is subject to a soft weight bounding to keep within the 0–1 range,

$$\Delta_{\text{sberr}} w_{ij} = [\Delta_{\text{err}}]_+ (1 - w_{ij}) + [\Delta_{\text{err}}]_- w_{ij} \qquad (9)$$

The two terms are then combined additively with a normalized mixing constant k_{hebb}:

$$\Delta w_{ij} = \epsilon [k_{\text{hebb}} (\Delta_{\text{hebb}}) + (1 - k_{\text{hebb}})(\Delta_{\text{sberr}})] \qquad (10)$$

PVLV Equations

See O'Reilly et al. (2007) for further details on the PVLV system. We assume that time is discretized into steps that correspond to environmental events (e.g., the presentation of a CS or US). All of the following equations operate on variables that are a function of the current time step t— we omit the t in the notation because it would be redundant. PVLV is composed of two systems, PV and LV, each of which in turn are composed of two subsystems (excitatory and inhibitory). Thus, there are four main value representation layers in PVLV (PV_e, PV_i, LV_e, LV_i), which then drive the dopamine layers (VTA/SNc).

Value Representations

The PVLV value layers use standard Leabra activation and kWTA dynamics, as described above, with the following modifications. They have a three-unit distributed representation of the scalar values they encode, where the units have preferred values of (0, .5, 1). The overall value represented by the layer is the weighted average of the

unit's activation times its preferred value, and this decoded average is displayed visually in the first unit in the layer. The activation function of these units is a "noisy" linear function (i.e., without the $x/(x + 1)$ nonlinearity to produce a linear value representation, but still convolved with Gaussian noise to soften the threshold, as for the standard units, Equation 4), with gain $\gamma = 220$, noise variance $\sigma = .01$, and a lower threshold $\Theta = .17$. The k for kWTA (average based) is 1, and the q value is .9 (instead of the default of .6). These values were obtained by optimizing the match for value represented with varying frequencies of 0–1 reinforcement (e.g., the value should be close to .4 when the layer is trained with 40% of 1 values and 60% of 0 values). Note that having different units for different values, instead of the typical use of a single unit with linear activations, allows much more complex mappings to be learned. For example, units representing high values can have completely different patterns of weights than those encoding low values, whereas a single unit is constrained by virtue of having one set of weights to have a monotonic mapping onto scalar values.

Learning Rules

The PV_e layer does not learn and is always just clamped to reflect any received reward value (r). By default, we use a value of 0 to reflect negative feedback, .50 for no feedback, and 1 for positive feedback (the scale is arbitrary). The PV_i layer units (y_j) are trained at every point in time to produce an expectation for the amount of reward that will be received at that time. In the minus phase of a given trial, the units settle to a distributed value representation based on sensory inputs. This results in unit activations y_j^-, and an overall weighted average value across these units denoted PV_i. In the plus phase, the unit activations (y_j^+) are clamped to represent the actual reward r (a.k.a., PV_e). The weights (w_{ij}) into each PV_i unit from sending units with plus-phase activations x_i^+ are updated using the delta rule between the two phases of PV_i unit activation states

$$\Delta w_{ij} = \epsilon (y_j^+ - y_j^-) x_i^+ \qquad (11)$$

This is equivalent to saying that the US/reward drives a pattern of activation over the PV_i units, which then learn to activate this pattern based on sensory inputs. In addition to the PV_e and PV_i layers, there is an additional PV_r layer that is associated with learning about reward detection. This system learns in the same way as the PV_i system but has a slower learning rate for weight decreases relative to increases.

The LV_e and LV_i layers learn in much the same way as the PV_i layer (Equation 11), except that the PV system filters the training of the LV values, such that they only learn from actual reward outcomes or when reward is

expected by the PV$_r$ system and not when no rewards are present or expected. This condition is as follows,

$$PV_{filter} = \min(PV_r, PV_i) < \theta_{min} \vee \max(PV_r, PV_i) > \theta_{max} \qquad (12)$$

$$\Delta w_i = \begin{cases} \epsilon\left(y_j^+ - y_j^-\right)x_i^+ & \text{if } PV_{filter} \\ 0 & \text{otherwise} \end{cases} \qquad (13)$$

where Θ_{min} is a lower threshold (0.20 by default), below which negative feedback is indicated and Θ_{max} is an upper threshold (0.80), above which positive feedback is indicated (otherwise, no feedback is indicated). Biologically, this filtering requires that the LV systems be driven directly by primary rewards (which is reasonable and is required by the basic learning rule anyway) and that they learn from dopamine dips driven by high PV$_r$ expectations of reward that are not met. The only difference between the LV$_e$ and LV$_i$ systems is the learning rate ϵ, which is .05 for LV$_e$ and .001 for LV$_i$. Thus, the inhibitory LV$_i$ system serves as a slowly integrating inhibitory cancellation mechanism for the rapidly adapting excitatory LV$_e$ system.

Finally, the NV layer signals stimulus novelty and produces dopamine bursts for novel stimuli, which slowly decay in magnitude as a stimulus becomes familiar. The habituation for this system is simply:

$$\Delta w_i = -NVx_i \qquad (14)$$

The PV, LV, and NV distributed value representations drive the dopamine layer (VTA/SNc) activations in terms of the difference between the excitatory and inhibitory terms for each. Thus, there is a PV delta, an LV delta, and an NV delta,

$$\delta_{pv} = PV_e - PV_i \qquad (15)$$

$$\delta_{lv} = LV_e - LV_i \qquad (16)$$

$$\delta_{nv} = NV \qquad (17)$$

The dopamine system integrates each of these inputs, using a temporal derivative computation to only produce brief bursts or dips relative to a baseline level of activation (this is the primary difference from the synaptic depression mechanism used in the earlier published version). The key issue is when to use each of the above values: If primary rewards are present or expected but not present, then the PV system dominates, and otherwise, LV + NV drive it. With the differences in learning rate between LV$_e$ (fast) and LV$_i$ (slow), the LV delta signal reflects recent deviations from expectations and not the raw expectations themselves, just as the PV delta reflects deviations from expectations about primary reward values. This is essen-

tial for learning to converge and stabilize when the network has mastered the task (as the results presented in this article show). These two delta signals need to be combined to provide an overall dopamine delta value, as reflected in the firing of the VTA and SNc units. One sensible way of doing so is to have the PV system dominate at the time of primary rewards, whereas the LV system dominates otherwise, by using the same PV-based filtering as holds in the LV learning rule:

$$\delta = \begin{cases} \left(\delta_{pv}^t - \delta_{pv}^{t-1}\right) & \text{if } PV_{filter} \\ \left(\delta_{lv}^t - \delta_{lv}^{t-1}\right) + \left(\delta_{nv}^t - \delta_{nv}^{t-1}\right) & \text{otherwise.} \end{cases} \qquad (18)$$

Special Basal Ganglia Mechanisms

Striatal Learning Function

Each stripe (group of units) in the striatum layer is divided into go versus no-go in an alternating fashion. The dopamine input from the SNc modulates these unit activations in the update phase by providing an extra excitatory current to go and an extra inhibitory current to the no-go units in proportion to the positive magnitude of the dopamine signal and vice versa for negative dopamine magnitude. This reflects the opposing influences of dopamine on these neurons (Frank, 2005; Gerfen, 2001). This updated phase of dopamine signal reflects the PVLV system's evaluation of PFC updates produced by gating signals in the plus phase. Learning on weights into the go/no-go units is based on the activation delta between the update (++) and plus phases,

$$\Delta w_i = \epsilon x_i(y^{++} - y^+) \qquad (19)$$

To reflect the finding that dopamine modulation has a contrast-enhancing function in the striatum (Frank, 2005; Nicola, Surmeier, & Malenka, 2000; Hernandez-Lopez, Bargas, Surmeier, Reyes, & Galarraga, 1997) and to produce more of a credit assignment effect in learning, the dopamine modulation is partially a function of the previous plus phase activation state,

$$g_e = \gamma[da]_+ y^+ + (1 - \gamma)[da]_+ \qquad (20)$$

where $0 < \gamma < 1$ controls the degree of contrast enhancement (.5 is used in all simulations), $[da]_+$ is the positive magnitude of the dopamine signal (0 if negative), $y+$ is the plus-phase unit activation, and g_e is the extra excitatory current produced by the da (for go units). A similar equation is used for extra inhibition (g_i) from negative da ($[da]_-$) for go units and vice versa for no-go units.

SNr and Thalamus Units

The SNr and thalamus (SNrThal) units provide a simplified version of the SNr/GPe/thalamus layers. They receive

a net input that reflects the normalized go/no-go activations in the corresponding striatum stripe,

$$\eta_j = \left[\frac{\sum \text{go} - \sum \text{no-go}}{\sum \text{go} + \sum \text{no-go}} \right]_+ \qquad (21)$$

(where $[]_+$ indicates that only the positive part is taken; when there is more no-go than go, the net input is 0). This net input then drives standard Leabra point neuron activation dynamics, with kWTA inhibitory competition dynamics that cause stripes to compete to update PFC. This dynamic is consistent with the notion that competition/selection takes place primarily in the smaller GP/SNr areas and not much in the much larger striatum (e.g., Mink, 1996; Jaeger, Kita, & Wilson, 1994). The resulting SNrThal activation then provides the gating update signal to PFC: If the corresponding SNrThal unit is active (above a minimum threshold; .1), then active maintenance currents in PFC are toggled.

This SNrThal activation also multiplies the per-stripe dopamine signal from the SNc,

$$\delta_j = \text{snr}_j \delta \qquad (22)$$

where snr_j is the snr unit's activation for stripe j and δ is the global dopamine signal (18).

Random Go Firing

The PBWM system only learns after go firing, so if it never fires go, it can never learn to improve performance. One simple solution is to induce go firing if a go has not fired after some threshold number of trials. However, this threshold would have to be either task specific or set very high, because it would effectively limit the maximum maintenance duration of PFC (because by updating PFC, the go firing results in loss of currently maintained information). Therefore, we have adopted a somewhat more sophisticated mechanism that keeps track of the average dopamine value present when each stripe fires a go,

$$\overline{\text{da}}_k = \overline{\text{da}}_k + \epsilon(\text{da}_k - \overline{\text{da}}_k) \qquad (23)$$

If this value is <0 and a stripe has not fired go within 10 trials, a random go firing is triggered with some probability (.1). We also compare the relative per-stripe dopamine averages, if the per-stripe dopamine average is low but is above zero and one stripe's $\overline{\text{da}}_k$ is .05, below the average of that of the other stripe's,

$$\text{if}(\overline{\text{da}}_k < .1) \text{ and } (\overline{\text{da}}_k - \langle \overline{\text{da}} \rangle < -.05); \text{go} \qquad (24)$$

a random go is triggered again with some probability (.1). Finally, we also fire random go in all stripes with some very low baseline probability (.0001) to encourage exploration.

When a random go fires, we set the SNrThal unit activation to be above go threshold, and we apply a positive dopamine signal to the corresponding striatal stripe so that it has an opportunity to learn to fire for this input pattern on its own in the future.

PFC Maintenance

PFC active maintenance is supported in part by excitatory ionic conductances that are toggled by go firing from the SNrThal layers. This is implemented with an extra excitatory ion channel in the basic V_m update equation (Equation 1). This channel has a conductance value of .5 when active. See Frank, Loughry, and O'Reilly (2001) for further discussion of this kind of maintenance mechanism, which has been proposed by several researchers, for example, Durstewitz, Seamans, and Sejnowski (2000), Gorelova and Yang (2000), Lewis and O'Donnell (2000), Dilmore, Gutkin, and Ermentrout (1999), Lisman, Fellous, and Wang (1999), and Wang (1999). The first opportunity to toggle PFC maintenance occurs at the end of the first plus phase and then again at the end of the second plus phase (third phase of settling). Thus, a complete update can be triggered by two gos in a row, and it is almost always the case that if a go fires for the first time, it will fire the next, because Striatum firing is primarily driven by sensory inputs, which remain constant.

Assessment of Selective Firing in BLA and OFC

We assessed whether a neuron in OFC or BLA that represented a particular US was more active after the presentation of the associated CS, relative to the presentation of the other CS.

The selectivity of firing in layer w equals the sum over the units k in layer w of the difference of activations x of a unit representing the US k in response to the associated CS i minus activation in response to nonassociated stimulus j, divided by the number of neurons in this layer: $sel_w = 1/n_w \times \Sigma_k(x_k i - x_k j)$.

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. R. (1995). Fear and the human amygdala. *Journal of Neuroscience, 15,* 5879–5891.

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.

Alexander, G., DeLong, M., & Strick, P. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9,* 357–381.

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

Berns, G. S., & Sejnowski, T. J. (1998). A computational model of how the basal ganglia produces sequences. *Journal of Cognitive Neuroscience, 10,* 108–121.

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.

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

Chudasama, Y., Kralik, J. D., & Murray, E. A. (2007). Rhesus monkeys with orbital prefrontal cortex lesions can learn to inhibit prepotent responses in the reversed reward contingency task. *Cerebral Cortex, 17,* 1154–1159.

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362,* 1585–1599.

Clarke, H. F., Robbins, T. W., & Roberts, A. C. (2008). Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *The Journal of Neuroscience, 28,* 10972–10982.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *The European Journal of Neuroscience, 18,* 1286–1294.

Damasio, A. R. (1994). *Descartes' error. Emotion, reason and the human brain.* New York: Avon Books.

Davis, M. (1992). The role of the amygdala in conditioned fear. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (1st ed., pp. 255–306). New York: Wiley-Liss.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature, 380,* 69.

Dilmore, J. G., Gutkin, B. G., & Ermentrout, G. B. (1999). Effects of dopaminergic modulation of persistent sodium currents on the excitability of prefrontal cortical neurons: A computational study. *Neurocomputing, 26,* 104–116.

Dominey, P., Arbib, M., & Joseph, J.-P. (1995). A model of corticostriatal plasticity for learning oculomotor associations and sequences. *Journal of Cognitive Neuroscience, 7,* 311–336.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience, 3(Suppl.),* 1184–1191.

Eagle, D. M., Baunez, C., Hutcheson, D. M., Lehmann, O., Shah, A. P., & Robbins, T. W. (2008). Stop-signal reaction-time task performance: Role of prefrontal cortex and subthalamic nucleus. *Cerebral Cortex, 18,* 178–188.

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex (New York, N.Y.: 1991), 10,* 308–317.

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

Fanselow, M. S., & Gale, G. D. (2003). The amygdala fear and memory. *Annals of the New York Academy of Sciences, 985,* 125–134.

Fanselow, M. S., & LeDoux, J. E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron, 23,* 229–232.

Featherstone, R. E., & McDonald, R. J. (2004). Dorsal striatum and stimulus-response learning: Lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a simple discrimination task. *Behavioural Brain Research, 150,* 15–23.

Ferrier, D. (1886). *Functions of the brain.* London: Smith and Elder.

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., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review, 113,* 300–326.

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

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

Gerfen, C. R. (2001). Molecular effects of dopamine on striatal-projection pathways. *Trends in Neurosciences, 23,* S64–S70.

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: Interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience, 115,* 1261–1279.

Gorelova, N. A., & Yang, C. R. (2000). Dopamine D1/D5 receptor activation modulates a persistent sodium current in rats prefrontal cortical neurons in vitro. *Journal of Neurophysiology, 84,* 75.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science (New York, N.Y.), 301,* 1104–1106.

Gray, T. S. (1999). Functional and anatomical relationships among the amygdala, basal forebrain, ventral striatum, and cortex. An integrative discussion. *Annals of the New York Academy of Sciences, 877,* 439–444.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia: I. A new functional anatomy. *Biological Cybernetics, 84,* 401–410.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 28,* 5623–5630.

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,* 105–118.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience and Biobehavioral Reviews, 34,* 701–720.

Hernandez-Lopez, S., Bargas, J., Surmeier, D. J., Reyes, A., & Galarraga, E. (1997). D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an l-type Ca^{2+} conductance. *Journal of Neuroscience, 17,* 3334–3342.

Holland, P. C., & Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Current Opinion in Neurobiology, 14,* 148–155.

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 J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 233–248). Cambridge, MA: MIT Press.

Houk, J. C., Bastianen, C., Fansler, D., Fishbach, A., Fraser, D., Reber, P. J., et al. (2007). Action selection and refinement in subcortical loops through basal ganglia and cerebellum. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362,* 1573–1583.

Houk, J. C., & Wise, S. P. (1995). Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: Their role in planning and controlling action. *Cerebral Cortex (New York, N.Y.: 1991), 5,* 95–110.

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,* 12921–12942.

Jaeger, D., Kita, H., & Wilson, C. J. (1994). Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *Journal of Neurophysiology, 72,* 2555–2558.

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,* 451–474.

Kita, T., Nishijo, H., Eifuku, S., Terasawa, K., & Ono, T. (1995). Place and contingency differential responses of monkey septal neurons during conditional place-object discrimination. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 15,* 1683.

LeDoux, J. (2000). Cognitive-emotional interactions: Listen to the brain. In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 129–156). New York: Oxford University Press.

Lewis, B. L., & O'Donnell, P. (2000). Ventral tegmental area afferents to the prefrontal cortex maintain membrane potential "up" states in pyramidal neurons via d(1) dopamine receptors. *Cerebral Cortex (New York, N.Y.: 1991), 10,* 1168–1175.

Lisman, J. E., Fellous, J. M., & Wang, X. J. (1999). A role for nmda-receptor channels in working memory. *Nature Neuroscience, 1,* 273–275.

McGeorge, A. J., & Faull, R. L. (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. *Neuroscience, 29,* 503–537.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia output and cognition: Evidence from anatomical, behavioral, and clinical studies. *Brain and Cognition, 42,* 183–200.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50,* 381–425.

Mishkin, M. (1964). Perseveration of central sets after frontal lesions in monkeys. In J. M. Warren & K. Abert (Eds.), *The frontal granular cortex and behavior* (pp. 219–241). New York: McGraw-Hill.

Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration and dissociation in early childhood. *Developmental Psychobiology, 40,* 255–265.

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends in Cognitive Sciences, 11,* 489–497.

Nicola, S. M., Surmeier, J., & Malenka, R. C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annual Review of Neuroscience, 23,* 185–215.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology, 191,* 507–520.

Noonan, M. P., Walton, M. E., Behrens, T. E. J., Sallet, J., Buckley, M. J., & Rushworth, M. F. S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 20547–20552.

Oja, E. (1983). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology, 15,* 267–273.

Ongür, D., Ferry, A., & Price, J. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *The Journal of Comparative Neurology, 460,* 425–449.

Ongür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex (New York, N.Y.: 1991), 10,* 206–219.

O'Reilly, R. (2006). Biologically based computational models of high-level cognition. *Science (New York, N.Y.), 314,* 91–94.

O'Reilly, R. C. (1996a). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation, 8,* 895–938.

O'Reilly, R. C. (1996b). *The Leabra model of neural interactions and learning in the neocortex.* PhD thesis, Carnegie Mellon University, Pittsburgh, PA.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation, 13,* 1199–1242.

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'Reilly, R. C., Mozer, M., Munakata, Y., & Miyake, A. (1999). Discrete representations in working memory: A hypothesis and computational investigations. *The Second International Conference on Cognitive Science* (pp. 183–188). Tokyo: Japanese Cognitive Science Society.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

Ostlund, S. B., & Balleine, B. W. (2007). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *Journal of Neuroscience, 27,* 4819–4825.

Palencia, C. A., & Ragozzino, M. E. (2005). The contribution of nmda receptors in the dorsolateral striatum to egocentric response learning. *Behavioral Neuroscience, 119,* 953–960.

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,* 865–870.

Pauli, W. M., Atallah, H. E., & O'Reilly, R. C. (2010). Integrating what & how/where with instrumental and Pavlovian learning: A biologically based computational model. In P. A. Frensch & R. Schwarzer (Eds.), *Cognition and neuropsychology - International perspectives on psychological science* (Vol. 1, pp. 71–95). East Sussex, UK: Psychology Press.

Pauli, W. M., Hazy, T. E., & O'Reilly, R. C. (2009). *Division of labor among multiple parallel cortico - basal ganglia - thalamic loops in Pavlovian and instrumental tasks: A biologically-based computational model.* Poster presented at the Multidisciplinary Symposium on Reinforcement Learning.

Postuma, R. B., & Dagher, A. (2006). Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. *Cerebral Cortex (New York, N.Y.), 16,* 1508–1521.

Preuss, T. M. (1995). Do rats have prefrontal cortex? The Rose–Woolsey–Akert program reconsidered. *Journal of Cognitive Neuroscience, 1,* 1–26.

Quirk, G. J., Repa, C., & LeDoux, J. E. (1995). Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: Parallel recordings in the freely behaving rat. *Neuron, 15,* 1029.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variation in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: Theory and research* (pp. 64–99). New York: Appleton-Century-Crofts.

Rolls, E. T. (1996). The orbitofrontal cortex. *Philosophical Transactions of the Royal Society of London, 351,* 1433–1444.

Rolls, E. T., Critchley, H. D., & Treves, A. (1997). Representation of olfactory information in the primate orbitofrontal cortex. *Journal of Neurophysiology, 75,* 1982.

Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: From affect to decision-making. *Progress in Neurobiology, 86,* 216–244.

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, N.Y.), 21,* 348–356.

Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences, 11,* 168–176.

Sacktor, T. C. (2011). How does PKM$_\zeta$ maintain long-term memory? *Nature Reviews, 12,* 9–15.

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

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

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., & Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron, 47,* 633–636.

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

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. (1993). 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.

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology, 74,* 1–57.

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.

Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science (New York, N.Y.), 321,* 848–851.

Stalnaker, T. A., Franz, T. M., Singh, T., & Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron, 54,* 51–58.

Stalnaker, T. A., Roesch, M. R., Franz, T. M., Burke, K. A., & Schoenbaum, G. (2006). Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making. *The European Journal of Neuroscience, 24,* 2643–2653.

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

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

Takahashi, Y. K., Roesch, M. R., Stalnaker, T. A., Haney, R. Z., Calu, D. J., Taylor, A. R., et al. (2009). The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron, 62,* 269–280.

Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). The orbitofrontal cortex: Neuronal activity in the behaving monkey. *Experimental Brain Research. Experimentelle Hirnforschung. Exprimentation Crbrale, 49,* 93–115.

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience, 30,* 31–56.

Wallis, J. D., Dias, R., Robbins, T. W., & Roberts, A. C. (2001). Dissociable contributions of the orbitofrontal and lateral prefrontal cortex of the marmoset to performance on a detour reaching task. *The European Journal of Neuroscience, 13,* 1797–1808.

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: The importance of nmda receptors to working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 19,* 9587.

Wickens, J. (1993). *A theory of the striatum.* Oxford, UK: Pergamon Press.

Wickens, J. R., Kotter, R., & Alexander, M. E. (1995). Effects of local connectivity on striatal function: Simulation and analysis of a model. *Synapse, 20,* 281–298.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *The European Journal of Neuroscience, 22,* 513–523.