Anatomy of a Decision: Striato-Orbitofrontal Interactions in Reinforcement Learning, Decision Making, and Reversal

Michael J. Frank and Eric D. Claus
University of Colorado at Boulder

The authors explore the division of labor between the basal ganglia–dopamine (BG-DA) system and the orbitofrontal cortex (OFC) in decision making. They show that a primitive neural network model of the BG-DA system slowly learns to make decisions on the basis of the relative probability of rewards but is not as sensitive to (a) recency or (b) the value of specific rewards. An augmented model that explores BG-OFC interactions is more successful at estimating the true expected value of decisions and is faster at switching behavior when reinforcement contingencies change. In the augmented model, OFC areas exert top-down control on the BG and premotor areas by representing reinforcement magnitudes in working memory. The model successfully captures patterns of behavior resulting from OFC damage in decision making, reversal learning, and devaluation paradigms and makes additional predictions for the underlying source of these deficits.

Keywords: decision making, neural network, basal ganglia, orbitofrontal cortex, reinforcement learning

What enables humans to make choices that lead to long-term gains, even when having to incur short-term losses? Such decision-making skills depend on the processes of action selection (choosing between one of several possible responses) and reinforcement learning (modifying the likelihood of selecting a given response on the basis of experienced consequences). Although all mammals can learn to associate their actions with consequences, humans are particularly advanced in their ability to flexibly modify the relative reinforcement values of alternative choices to select the most adaptive behavior in a particular behavioral, spatial, and temporal context.

The behavioral and cognitive neurosciences have identified two neural systems that are involved in such adaptive behavior. On the one hand, the basal ganglia (BG) and the neuromodulator dopamine (DA) are thought to participate in both action selection and reinforcement learning (Beiser & Houk, 1998; Brown, Bullock, & Grossberg, 1999, 2004; Frank, 2005; Frank, Loughry, & O'Reilly, 2001; Gurney, Prescott, & Redgrave, 2001; Mink, 1996; O'Reilly & Frank, 2006). Patients with Parkinson's disease (PD), who have low levels of DA in the BG, are impaired at making choices that require learning from trial and error (Cools, 2005; Knowlton, Mangels, & Squire, 1996; Shohamy et al., 2004). Biologically based computational models demonstrate how the BG-DA system can learn to make adaptive choices (Brown et al., 2004; Frank,

2005) and provide an account for how this is impaired in PD (Frank, 2005).

On the other hand, various lines of evidence suggest that ventromedial and orbitofrontal cortices are critical for adaptive decision making in humans and that homologous areas support more primitive forms of this behavior in animals (Kringelbach & Rolls, 2004; Rolls, 1996; Schoenbaum, Setlow, Saddoris, & Gallagher, 2003; Tremblay & Schultz, 2000). Patients with orbitofrontal cortex (OFC) damage exhibit decision-making deficits in their everyday lives, which have also been documented in the laboratory (Bechara, Damasio, Tranel, & Anderson, 1998). Drug abusers, who are almost by definition poor decision makers, have reduced OFC metabolism and gray matter volume (Milham et al., 2006; Volkow, Fowler, & Wang, 2003). Finally, OFC lesions impair one's ability to learn when previous reward associations no longer apply, as in reversal learning (Chudasama & Robbins, 2003; Jones & Mishkin, 1972). Thus, both the BG-DA and OFC systems have been implicated in decision making and reinforcement and reversal learning, but surprisingly little theoretical work addresses how these systems are related or interact. Given that the OFC is a recent structure phylogenetically, a reasonable question to ask is "What unique function does the OFC contribute to decision making that is not supported by the more primitive BG-DA system?"

In this article, we extend a previous neural network model of the BG-DA system (Frank, 2005) to explore additional contributions of the OFC that enable adaptive and flexible decision making. In brief, this account is consistent with the idea that the BG system is specialized to slowly integrate positive and negative outcomes over multiple trials, resulting in the ingraining of motor habits (Jog, Kubota, Connolly, Hillegaart, & Graybiel, 1999). The model accomplishes this by learning *go* to facilitate responses that generally lead to positive outcomes while concurrently learning *no-go* to suppress inappropriate responses (Frank, 2005). In contrast, the prefrontal cortex (PFC) actively maintains information in working memory via persistent neural firing (Fuster, 1997; Goldman-Rakic, 1995; Miller, Erickson, & Desimone, 1996), and this has a top-

down biasing effect to guide behavior (J. D. Cohen, Dunbar, & McClelland, 1990; Miller & Cohen, 2001). This active incorporation of recent contextual information can both complement and compete with more latent habitual representations (e.g., Morton & Munakata, 2002). Further, this model adheres to a dominant theory of PFC function, suggesting that separate areas within the PFC are distinguished primarily by the type of information maintained in working memory (Goldman-Rakic, 1995), with the OFC representing reward values (Elliott, Dolan, & Frith, 2000; Schoenbaum & Roesch, 2005; Schoenbaum & Setlow, 2001; Tremblay & Schultz, 1999). More specifically, medial and lateral areas of the OFC represent positive and negative outcomes of decisions, respectively (Gottfried, O'Doherty, & Dolan, 2002; O'Doherty, Kringelback, Rolls, Hornak, & Andrews, 2001; Rolls & Kringelbach, 2003; Ursu & Carter, 2005). By actively maintaining this information in working memory, these OFC areas have a top-down biasing effect on response selection processes of the BG and premotor cortex.

The net result is that the BG-DA system is well suited to learn to make choices based on their relative *probability* of resulting in a positive outcome, but the OFC is necessary to provide active working memory of the relative *magnitudes* of gain–loss information. The combination of the two systems results in an improved representation of the true expected value of a given decision. Through explicit simulations, our augmented model accounts for decision-making deficits in OFC patients, making specific predictions for different patterns of lateral versus medial lesions. Further, the same OFC mechanisms are leveraged to be able to quickly modify reward associations when these suddenly change, as in reversal learning. Finally, simulation of OFC and striatal damage in our model produces qualitatively similar behavioral deficits observed in these clinical populations.

Because our primary goal is to develop a theoretical framework for understanding the differential neural system contributions to decision making, we review studies with relevant neural data across rats and nonhuman primates (for which direct neural recording and lesion data are abundant) and humans (providing assurance that similar neuronal functions hold across species). These converging data constrain the mechanisms simulated in our model, so that it is not dependent on a single result from any particular study but rather includes results that have been observed across multiple studies and species.

Anatomy of a Decision

Several interacting brain regions participate in decision-making processes, including but not limited to the BG, parietal cortex, PFC, and subdivisions therein, and both the anterior and posterior cingulate cortices (McCoy & Platt, 2005; Platt, 2002). Here we focus on (a) the subcortical BG-DA system and its interactions with the premotor cortex and (b) the ventromedial and orbitofrontal cortices and their top-down influence on (a).

BG-DA

Our framework for conceptualizing the role of the BG in decision making builds on suggestions by Mink and others (O. Hikosaka, 1998; Mink, 1996) about the role of the same structures in motor control. In particular, the BG are thought to modulate the selection of actions being considered in the frontal cortex (Brown

et al., 2004; Frank, 2005; Frank et al., 2001; Gurney et al., 2001; O. Hikosaka, 1998; Mink, 1996; Rubchinsky, Kopell, & Sigvardt, 2003). More specifically, two main projection pathways from the striatum proceed through different BG output structures on the way to the thalamus and up to the cortex (Alexander & Crutcher, 1990a; Alexander, Crutcher, & DeLong, 1990; see Figure 1a). Activity in the direct pathway sends a go signal to facilitate the execution of the most appropriate cortical response, whereas activity in the indirect pathway sends a no-go signal to suppress competing responses.

We do not intend to suggest that the BG act as a homunculus, somehow magically pulling the levers to decide which action is most appropriate. Rather, our account of how the BG learn the distinction between adaptive and maladaptive responses incorporates findings by Schultz and others that phasic changes in DA release are critically involved in reinforcement learning processes that can train the BG to function adaptively (e.g., Schultz, 1998, 2002; Schultz, Dayan, & Montague, 1997; Wickens, 1997). Under normal conditions, DA cells fire at intrinsic baseline levels. When animals (both rats and primates) make choices that lead to unexpected rewards (i.e., rewards that were not predicted on the basis of previous experience with the chosen option), transient bursting of dopaminergic cells and DA release are observed. Conversely, choices that do not lead to reward when a reward is expected are associated with dips in DA firing that drop below baseline (Hollerman & Schultz, 1998; Pan, Schmidt, Wickens, & Hyland, 2005; Satoh, Nakai, Sato, & Kimura, 2003; Schultz, 2002).

Computational models have suggested that the primary effect of DA in the BG is to enhance go firing and suppress no-go firing (Brown et al., 2004; Frank, 2005). This is supported by the observation that go and no-go cells primarily express D_1 and D_2 receptors, respectively (Aubert, Ghorayeb, Normand, & Bloch, 2000; Gerfen, 1992; Wise, Murray, & Gerfen, 1996). Given that DA is excitatory to synaptic input on D_1 receptors (Hernandez-Lopez, Bargas, Surmeier, Reyes, & Galarraga, 1997), its effect is to increase go activity. And given that DA is inhibitory on D_2 receptors (Hernandez-Lopez et al., 2000), its effect is to suppress no-go activity. Thus, increases in DA during positive reinforcement lead to transient changes in go/no-go activity, which in turn drive Hebbian learning in opposite directions in the two types of cells so that the animal is more likely to facilitate the action that led to reinforcement (Frank, 2005; Houk & Wise, 1995; Joel & Weiner, 1999). This is consistent with observations that DA enhances synaptic plasticity and promotes long-term potentiation via D_1 receptors (go pathway) while promoting long-term depression via D_2 receptors (no-go; Centonze, Picconi, Gubellini, Bernardi, & Calabresi, 2001; Nishi, Snyder, & Greengard, 1997). Conversely, DA dips during negative reinforcement may also be adaptive, in that they can drive no-go learning to avoid selecting the nonreinforced response in the future (Frank, 2005).[1] Specifically, low

---

[1] Although the change in firing rate associated with DA dips is smaller than that of the bursts (due to already low baseline firing rates of DA cells; Bayer & Glimcher, 2005), this asymmetry does not mean that dips are less effective in driving learning. For example, the smaller range of DA dips is likely compensated for by a counteracting asymmetry in the receptor sensitivity to dips versus bursts. In particular, because DA has greater affinity for the D_2 than for the D_1 class of receptors (Creese et al., 1983), D_2 receptors are very sensitive to low tonic DA levels, whereas large increases in DA are required to functionally stimulate D_1 receptors (e.g.,
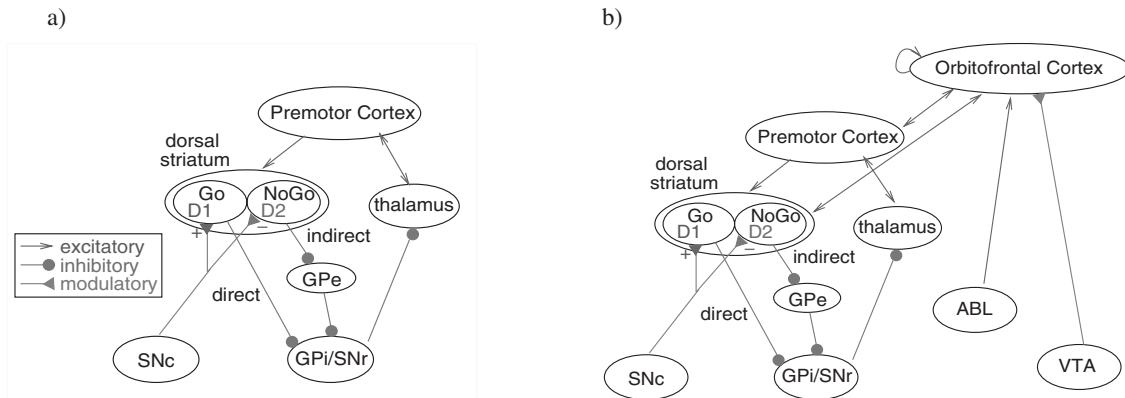
Figure 1. a: The cortico-striato-thalamo-cortical loops, including the direct and indirect pathways of the basal ganglia. The cells of the striatum are divided into two subclasses based on differences in biochemistry and efferent projections. The go cells project directly to the the internal segment of the globus pallidus (GPi)–substantia nigra pars reticulate (SNr), and their activity disinhibits the thalamus, thereby facilitating the execution of a cortical response. The no-go cells are part of the indirect pathway to the GPi-SNr and have an opposing effect, suppressing actions from being executed. Dopamine from the substantia nigra pars compacta (SNc) projects to the dorsal striatum, differentially modulating activity in the direct and indirect pathways by activating different receptors: The go cells express the D_1 receptor, and the no-go cells express the D_2 receptor. b: The same circuit with additional influence from the orbitofrontal cortex, which can maintain reinforcement-related information in working memory and provide top-down biasing on the more primitive basal ganglia system, in addition to direct influencing of response selection processes in the premotor cortex. The orbitofrontal cortex receives information about the relative magnitude of reinforcement values from the basolateral nucleus of the amygdala (ABL), which it can also maintain in working memory. Dopamine from the ventral tegmental area (VTA) projects to the ventral striatum (not shown) and the orbitofrontal cortex. GPe = external segment of the globus pallidus.

levels of DA release no-go cells from inhibition (via lack of binding to D_2 receptors), allowing them to become more excited than their go counterparts and driving Hebbian learning in the opposite direction of DA bursts. Supporting this hypothesis, blockade of D_2 receptors is associated with enhanced no-go activity and associated increases in long-term potentiation (Calabresi et al., 1997; Finch, 1999; Robertson, Vincent, & Fibiger, 1992).

As DA bursts and dips reinforce go and no-go representations in the BG, the previous model showed that the most adaptive (i.e., rewarding) responses represented in premotor areas tend to be facilitated, whereas less adaptive ones are suppressed. Further, as the BG learn to facilitate adaptive responses, the associated adaptive representations may become enhanced in premotor cortical areas. In this way, DA reward processes within the BG may ingrain prepotent motor habits in cortical areas (Frank, 2005). Once these habits are ingrained, there is less need for selective facilitation by the BG. This is consistent with observations that dopaminergic integrity within the BG is much more critical for the acquisition rather than the execution of instrumental responses (Parkinson et al., 2002; Smith-Roe & Kelley, 2000) and with recent physiological observations that learning-related activity is initially seen in the BG and is observed only later in the frontal

cortex (Delgado, Miller, Inati, & Phelps, 2005; Pasupathy & Miller, 2005).

OFC

Several lines of evidence suggest that the OFC is critical for more flexible, advanced forms of representing expected value (Rolls, 2004; Schoenbaum & Roesch, 2005; Tremblay & Schultz, 1999, 2000). In particular, we suggest that the OFC participates in decision making much in the same way as the dorsolateral PFC is thought to subserve cognitive control processes, via maintenance of goal-directed actions (Frank et al., 2001; Miller & Cohen, 2001; O'Reilly, Noelle, Braver, & Cohen, 2002; Schoenbaum & Setlow, 2001). Although dissociations have been found between the dorsolateral PFC and the OFC for the processes of working memory and decision making, respectively (Bechara et al., 1998), we suggest that the difference in these two regions is not in function but in the content of working memory representations (e.g., Goldman-Rakic, 1995). Specifically, we argue that the OFC is specialized to maintain recent gain–loss information in working memory and that this can have a top-down biasing effect on the more primitive BG system, as well as having direct effects on motor responding in frontal motor areas (Wallis & Miller, 2003). The firing of OFC neurons is thought to represent the current behavioral strategy; this is particularly evident when the outcome predicted by a stimulus is altered or reversed (Schoenbaum, Chiba, & Gallagher, 2000). The need for such a system in decision making would be particularly evident in cases for which the best decision depends on rapidly changing factors along with consid-

Goto & Grace, 2005). Thus, smaller decreases in DA firing may be sufficient to functionally affect D_2 receptors, which drives no-go learning in our model. Further, recent evidence shows that information within DA dips is carried by the duration of these dips rather than their magnitudes (Bayer, 2004).

erations of relative magnitudes of gains and losses, rather than simply choosing on the basis of the usual outcome of a particular decision.

Medial and lateral OFC functional distinctions. We suggest that the OFC represents both positive and negative reinforcement values. Although in principle these representations need not be separated anatomically, several investigations have provided evidence for a functional distinction between medial and lateral subregions within the OFC, complementing previous anatomical distinctions (Carmichael & Price, 1995; Cavada, Company, Tejedor, Cruz-Rizzolo, & Reinoso-Suarez, 2000). First, Iversen and Mishkin (1970) found that primates with either medial or lateral OFC lesions were impaired in a reversal learning paradigm and that there was a dissociable pattern of deficits. Specifically, medial lesions impaired the ability to learn to associate previously non-rewarding stimuli with reward, whereas lateral lesions resulted in failure to withhold responding to previously rewarding stimuli. Neuroimaging studies have since shown that the medial OFC becomes active when participants learn to choose rewarding stimuli, whereas the lateral OFC is more active when learning about punishing stimuli (O'Doherty et al., 2001). Further, the amplitude of the activity within the medial or lateral OFC significantly correlated with the magnitude of the reward or punishment, respectively. A correlation has also been found between medial OFC activity and subjective ratings of odor pleasantness, whereas lateral activity was correlated with unpleasantness (Rolls & Kringelbach, 2003). This anatomical distinction also seems to apply to neutral stimuli that predict positive and negative valence. For example, medial OFC activation was observed when neutral visual stimuli (faces) were paired with pleasant odors, whereas the lateral OFC became active during association with unpleasant odors (Gottfried et al., 2002). Finally, the same distinction has been observed during the retrieval of autobiographical memories, with happy memories activating the medial OFC and sad memories activating the lateral OFC (Markowitsch, Vandekerckhovel, Lanfermann, & Russ, 2003). On the basis of much of this evidence, it has been suggested that the medial OFC monitors the association between stimuli and correct or rewarded responses, whereas the lateral OFC overrides choices based on previous but no longer applicable reward values (Elliott et al., 2000).

The OFC receives reinforcement information from the amygdala. How does the OFC, a higher order frontal cortical structure, encode information about reinforcement? Substantial evidence suggests that the ability of the OFC to predict outcome values depends on input from the basolateral nucleus of the amygdala (ABL; Holland & Gallagher, 2004). First, experimental evidence shows that both negative (aversive) and positive (rewarding) outcomes are encoded in the ABL (Baxter & Murray, 2002). Whereas the ABL and OFC each receive sensory information from sensory cortical areas, the OFC is thought to learn the affective nature of sensory information via dense projections from the ABL (Cavada et al., 2000). Single-cell recording demonstrates that ABL lesions prevent OFC neurons from responding to expected outcomes (Schoenbaum et al., 2003), and disconnection of the ABL from the OFC impairs control of response selection by reinforcer value (Baxter, Parker, Lindner, Izquierdo, & Murray, 2000). Complementary roles have been proposed for the ABL and OFC, such that the ABL acquires associations between cues and outcomes, whereas the OFC subsequently maintains these associations in memory, updates them with newly experienced information, and

uses them to guide behavior (Pickens et al., 2003). Similarly, Rolls (2004) suggested that the OFC takes over some of the role of the amygdala in primates, enabling faster adaptation to changing reward values to implement behavioral control. In other words, whereas the amygdala may require multiple trials with a negative outcome before reversing its bias on responses, the OFC is able to maintain outcome information from the amygdala and then use this information to bias response processes on the next trial.

The OFC maintains reinforcement information in working memory. The notion that the OFC maintains working memory for reward information is well supported. For example, single-cell recordings in rats showed that neurons in the OFC reliably fire once a behavioral strategy is adopted, suggesting that the OFC biases a motor response on the basis of reward information currently being held online (Schoenbaum, Chiba, & Gallagher, 1999). In primates, single-cell OFC recordings show increased firing during a delay period for cues that predict reward (K. Hikosaka & Watanabe, 2000), and the magnitude of this firing is correlated with predicted reward magnitude (Roesch & Olson, 2004). Reward-related delay period activity has also been observed in the OFC, where activity in some cells increases to predictors of positive reinforcement, whereas other cells respond to negative reinforcement, with some evidence for a medial–lateral distinction (Rosenkilde, Bauer, & Fuster, 1981).

Differential responding has also been observed in the primate OFC for short- and long-range reward expectancy (K. Hikosaka & Watanabe, 2004), such that some neurons coded for rewards expected within the current trial, whereas others fired in anticipation of a preferred reward that would occur within the next four trials (a 50-s duration). Further, lesions to the OFC in rats resulted in impulsive choices in a delay discounting paradigm, such that lesioned animals were more likely to choose a smaller immediate reward over a larger delayed reward, over delay periods that ranged from 1 s to 30 s (Mobini et al., 2002). These results are consistent with the idea that the OFC is required to maintain reward associations in working memory over time and for representing differential reward magnitudes. Finally, in humans, patients with OFC damage failed to learn to choose advantageously in a gambling task. Specifically, these patients not only made choices that led to long-term losses but also failed to show increased skin conductance response associated with the emotional value of choices in control participants (Bechara, Tranel, Damasio, & Damasio, 1996). In addition, many of these patients could verbalize whether a deck of cards was good or bad but did not appropriately apply this information when performing the task, suggesting that they were unable to hold reward information online to bias an appropriate motor response.

Top-down OFC effects on the BG and adaptive responding. Given that the OFC is important for encoding and holding online reward-related information, it is reasonable to explore how it interacts with the lower level reinforcement-learning BG system. First, it is important to consider anatomical projections between the OFC and BG (Haber, Kunishio, Mizobuchi, & Lynd-Balta, 1995; Kemp & Powell, 1970). As mentioned above, the OFC is part of a BG circuit that also includes the ventral striatum, globus pallidus, and the dorsomedial nucleus of the thalamus (Alexander et al., 1990). The medial OFC projects to the medial ventral striatum, whereas the lateral OFC projects more densely to the core of the nucleus accumbens (NAc; Haber et al., 1995). Of interest,

Haber et al. (1995) suggested that neurons within the striatum may be influenced by disparate parts of the OFC.

There is also functional evidence for interactions between the OFC and BG. Like the OFC, striatal neurons reverse their reward-related responses in reversal learning paradigms (Setlow, Schoenbaum, & Gallagher, 2003; Watanabe & Hikosaka, 2005). Activity in ventral caudate and striatal areas during reversal is thought to reflect top-down activity from the OFC (Rolls, 1999). In humans, concurrent ventral striatal and orbitofrontal activation were shown to coincide with reversal behavior in a probabilistic reversal paradigm (Cools, Clark, Owen, & Robbins, 2002). Further, medial OFC activity was associated with maintenance of the current behavioral strategy activated, whereas lateral OFC and ventral striatal activity predicted a switch in behavior (O'Doherty, Critchley, Deichmann, & Dolan, 2003). Finally, ventral striatal lesions in monkeys led to reversal learning impairments (Annett, McGregor, & Robbins, 1989; Rolls, 1999; Schoenbaum & Setlow, 2003), and medication caused these same deficits in human patients with PD, which are thought to stem from an "overdose" of DA in the ventral striatum (Cools, Barker, Sahakian, & Robbins, 2001; Swainson, Rogers, Sahakian, Summers, Polkey, & Robbins, 2000).

The above evidence suggests that the BG and OFC intimately interact in reinforcement learning and decision making. Given that medial OFC areas tend to be activated by positive outcomes and maintenance of current strategies, it is plausible that these areas have a top-down biasing effect on go responding within the BG. Conversely, lateral OFC areas respond to negative events and are predictive of a switch in behavior, supporting the possibility that they support no-go BG responding (see the Discussion section for possible additional interactions with the anterior cingulate [ACC]). Next, we formally explore this division of labor between the systems via explicit computational simulations to determine whether their emergent dynamics can learn to make decisions based on their true expected value (e.g., a combination of frequency and magnitude of gains and losses).

Neural Network Simulations

All models described in this article are implemented in the Leabra framework, which combines Hebbian learning and a biologically plausible version of error-driven learning, together with inhibitory competition, into one coherent framework (see the Appendix; see also O'Reilly, 1998; O'Reilly & Munakata, 2000).

We begin with a short description of the previous model of striatocortical circuits in decision making (Figure 2; Frank, 2005). This model learns to make choices that are likely to lead to positive
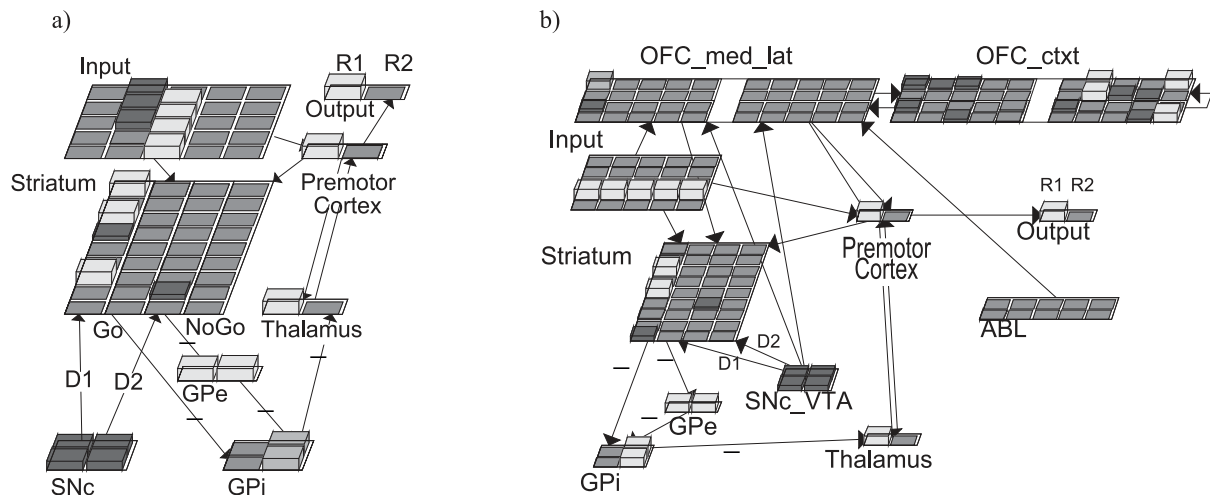

Figure 2. a: The Frank (2005) neural network model of the primitive basal ganglia–dopamine (BG-DA) system depicted in Figure 1a. Rectangles represent units, with height and color reflecting neural activity. The premotor cortex selects an output response via direct projections from the sensory input and is modulated by the BG projections from thalamus. Go units are in the left half of the striatum layer, and no-go units are in the right half, with separate columns for the two responses (R1, Response 1; R2, Response 2). In the case shown, the striatum go is stronger than the no-go for R1, inhibiting the internal segment of the globus pallidus (GPi), disinhibiting the thalamus, and facilitating execution of the response in the cortex. A tonic level of DA is shown in the substantia nigra pars compacta (SNc); a burst or dip ensues in a subsequent error feedback phase (not shown), causing corresponding changes in go/no-go unit activations, which drive learning. This model learns to select responses that are more likely to result in positive than in negative reinforcement. b: The augmented model, including orbitofrontal contributions. The BG circuitry is as in panel a. The orbitofrontal cortex (OFC) is modeled in two separate layers, both of which have medial and lateral subdivisions. In OFC_med_lat, a distributed pattern of activity in the left-hand side reflects a medial representation about previously positive associations for the input stimulus. This activity has a top-down effect on responding within the striatum, as well as an effect directly on the premotor cortex. OFC_ctxt is a context layer that maintains recent reinforcement information in working memory and biases activity in OFC_med_lat for use in behavioral decisions. The lateral representation in the context layer depicts memory for a negative outcome of an alternate stimulus in a previous trial but does not influence OFC_med_lat or behavior in this trial (see text for details). ABL = basolateral nucleus of the amygdala; GPe = external segment of the globus pallidus; VTA = ventral tegmental area. SNc_VTA = combined layer representing DA cells in both SNc and VTA.

reinforcement while suppressing those that have overall negative associations. We then show that this model fails to make adaptive decisions in a gambling task in which good choices are associated with a low probability of large gains and a high probability of small losses, whereas bad choices are associated with a high probability of small gains and a low probability of large losses. Thus, this model does not optimize choices on the basis of expected values of alternative decisions but largely capitalizes only on frequencies. We then demonstrate that an augmented model that includes medial and lateral areas of the OFC and the amygdala can successfully solve this difficult decision-making task. The same model is also shown to outperform the primitive model in a reversal learning task, providing a unified framework for OFC involvement in complex decision making (gambling tasks) and reversal learning. Next, we apply the models to a devaluation procedure to examine the contributions of the different areas in habit learning versus cognitive action–outcome expectancy and show that lesions of the striatum or OFC produce results qualitatively similar to those observed in animals with the two types of lesions. Finally, we use the models to explain the classic framing effect reported in the behavioral decision-making literature, by leveraging the medial–lateral distinction in the OFC to show how risk aversion for gains and risk seeking for losses may be accounted for at the neural level.

Primitive BG-DA Model

The details of the BG model are described in Frank (2005). In brief, the premotor cortex represents and considers two possible responses (R1 and R2) for each input stimulus. The BG system modulates which one of these responses is facilitated and which is suppressed by signaling go or no-go to each of the responses. The four columns of units in the striatum represent, from left to right, go-R1, go-R2, no-go-R1 and no-go-R2. Go and no-go representations for each response compete at the level of the internal segment of the globus pallidus, such that stronger go representations lead to disinhibition of the corresponding column of the thalamus, which in turn amplifies and facilitates the execution of that response in the premotor cortex. Concurrently, the alternative response is suppressed.

Striatal go/no-go representations are learned via phasic changes in simulated DA firing in the substantia nigra pars compacta layer during positive and negative reinforcement. After correct responses, increases in DA firing excite go units for the response just selected while suppressing no-go units, via simulated D_1 and D_2 receptors. Conversely, decreases in DA after incorrect responses result in increased no-go activity for that response. This DA modulation of go/no-go activity drives learning as described above.

This model can successfully learn to make choices that are more often reinforced than not in challenging probabilistic tasks, discriminating between subtle differences in reinforcement value (Frank, 2005). Further, this model has shown that both the go and no-go learning mechanisms are necessary to learn subtle discriminations between reinforcement values of decisions. Networks with the no-go pathway disconnected were impaired at learning probabilistic discriminations relative to networks that could compare both go and no-go learning for multiple possible responses. More recent simulations explored additional contributions of the subthalamic nucleus in the overall BG circuitry, showing that this brain area can dynamically modulate the threshold for executing responses depending on the degree of response conflict present (Frank, in press). This model also showed that BG network dynamics during response selection are consistent with available data in both intact and PD states.

Empirical confirmation of the model's mechanisms comes from behavioral studies showing that differing levels of BG DA are associated with differential go and no-go learning from positive versus negative reinforcement in medicated and nonmedicated PD patients (see Figure 3; Frank, Seeberger, & O'Reilly, 2004). Similar predictions were borne out in young healthy participants taking dopaminergic medications (Frank & O'Reilly, in press) and in electrophysiological studies of individual differences in reinforcement learning (Frank, Woroch, & Curran, 2005).

Despite this positive evidence, the BG model as it stands is not adequate to account for more complex real-world decisions. In particular, it is not equipped to incorporate relative magnitudes of gains and losses together with their probability of occurring to choose the most adaptive option. As shown below, this is because there is not sufficient information in the DA signal (or in its postsynaptic effects) to allow graded differences in gain–loss magnitudes to significantly affect striatal learning. Although the size of DA bursts was recently reported to scale with increasing reward magnitude (Tobler, Fiorillo, & Schultz, 2005), we hypothesize that a single large DA burst (or dip) may not be sufficient to counteract the learning effects of several smaller DA dips (or bursts). In contrast, a differential DA magnitude effect may have substantial effects on frontal activity, because the effects of a single large DA burst persist over time (Seamans & Yang, 2004) as reward representations are maintained in working memory. The simulations reported below more strongly support this notion.

Augmenting the Model: Orbitofrontal Contributions

In the augmented BG-OFC model (see Figure 2b), the BG-DA system is the same as in the prior primitive model and continues to integrate positive and negative decision outcomes via go and no-go learning. The OFC is divided into two layers, OFC_med_lat and OFC_ctxt, each of which has corresponding medial and lateral subdivisions displayed on the left- and right-hand sides, respectively. The OFC_med_lat layer represents current positive/negative outcome expectancy and has top-down biasing effects on response selection processes in both the premotor cortex and striatum. The OFC_ctxt layer represents contextual information about previously experienced gains and losses by maintaining these in working memory. The division of the OFC by function into two layers is consistent with observations that different parts of the OFC are activated during behavioral control and memory for reinforcement outcome (O'Doherty et al., 2003). Alternatively, the two layers could reflect laminar divisions within a region of the OFC, with OFC_ctxt representing superficial cortical layers (2 and 3), and OFC_med_lat representing output projection neurons in deep layers (5 and 6).

The two OFC layers are interconnected in a one-to-one fashion, such that current expectations about a particular stimulus in OFC_med_lat are influenced by working memory about their previous outcomes maintained in OFC_ctxt. When positive or negative outcomes are actually experienced following a decision, they reciprocally update working memory in OFC_ctxt on the subsequent trial. Recurrent connections in OFC_ctxt result in
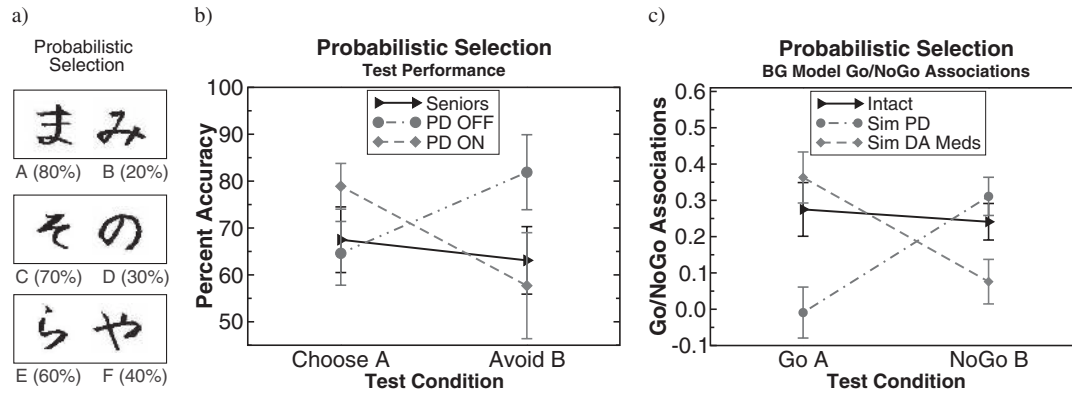
Figure 3. Empirical support for a neural network model of the basal ganglia–dopamine (BG-DA) system. a: Example stimulus pairs used in a cognitive probabilistic learning task designed to minimize verbal encoding. One pair is presented per trial, and the participant makes a forced choice. The frequency of positive feedback for each choice is shown. b: Performance in patients with Parkinson's disease (PD) on and off medication in a test phase in which Stimuli A and B were re-paired with more neutral stimuli (Frank et al., 2004). Patients on medication more reliably chose Stimulus A (which depends on having learned from positive feedback) than avoided Stimulus B (which depends on negative feedback learning). Patients off medication showed the opposite pattern. Findings for Seniors not addressed in this article. c: This pattern of results was predicted by the Frank (2005) model. The figure shows go/no-go associations for Stimulus A and no-go/go associations for Stimulus B, recorded from the model's striatum after having been trained on the same task used with patients off and on DA medication (Sim PD and Sim DA Meds, respectively). Error bars reflect standard error across 25 runs of the model with random initial weights.

persistent maintenance of recent stimulus–reinforcement-related information. There is some hysteresis in OFC_ctxt, so that working memory representations decay over time, but as a result of nonlinear activation dynamics, large magnitude gains–losses tend to be less subject to this decay than are smaller gains–losses. The parameters chosen for the decay of working memory activity (see the Appendix for details) are generally consistent with available neurophysiological data (Major & Tank, 2004) and other biologically detailed computational models of persisting PFC activity over delays (Durstewitz, Seamans, & Sejnowski, 2000), both showing that PFC maintenance activity is robust over tens of seconds.[2]

Note that OFC stimulus–reinforcement representations are distributed and learned. The OFC_med_lat layer receives direct sensory projections from the input layer, with random initial weights. Thus, the first time the network encounters a particular input stimulus, it will randomly activate a subset of medial and lateral OFC neurons. Depending on the outcome of the model's response to that stimulus, either the medial or the lateral associations will be strengthened (see Figure 4). In any subsequent exposure to that stimulus, the OFC representation is determined by a combination of input projections and projections from the OFC_ctxt. In this manner, the currently relevant stimulus is activated in OFC_med_lat in conjunction with its contextual influences. Contextual memory for reinforcement values of other stimuli not currently present in the input tend to be suppressed via local inhibitory competition (as they have to compete with representations having additional activation from sensory input) and therefore do not influence responding in that trial.

Outcome magnitude from the ABL is represented in the OFC. Reinforcement magnitude information arrives to the OFC from the ABL layer, where gain–loss magnitude is encoded in proportion to overall activity levels. For simplicity, we apply these magnitudes

directly onto the ABL layer during positive and negative outcomes; this work does not address how the ABL comes to represent magnitude information but simply explores the magnitudes' effects on the OFC and striatal function. Greater magnitudes in the ABL have a stronger effect on OFC activity, as is observed empirically (see above).

The balance between medial and lateral OFC representations during positive or negative reinforcement is determined by two factors. First, the weights between the ABL units and medial–lateral OFC units are dynamically adjusted depending on whether the reinforcement is positive or negative.[3] Second, increases in DA firing during positive outcomes activate medial representations and deactivate lateral representations. This feature of the model was motivated by the dense projections from midbrain DA cells to medial areas of the PFC (Lindvall, Bjorklund, & Divac, 1978) and by observations that DA increases excitability of synaptically

---

[2] This decay activity refers to the passive decay of neural activity in the PFC. The length of persistence of activity can theoretically be shorter depending on subsequent input (i.e., if the contents of working memory are updated; Frank et al., 2001; O'Reilly & Frank, 2006) or longer (if the working memory representations are refreshed because they continue to be task relevant).

[3] During positive outcomes, weights are zeroed out between the ABL and the lateral OFC, whereas during negative outcomes, weights are zeroed out for connections between the ABL and the medial OFC. This is an implementational detail that was motivated by the fact that the ABL seems to activate to both positive and negative magnitudes in imaging studies (Anderson et al., 2003). However, in reality, there are likely different populations of cells within the ABL that represent positive and negative valence, and these could potentially have differential projections to the medial and lateral OFC. The present implementation is therefore not critical to the functioning of the network.
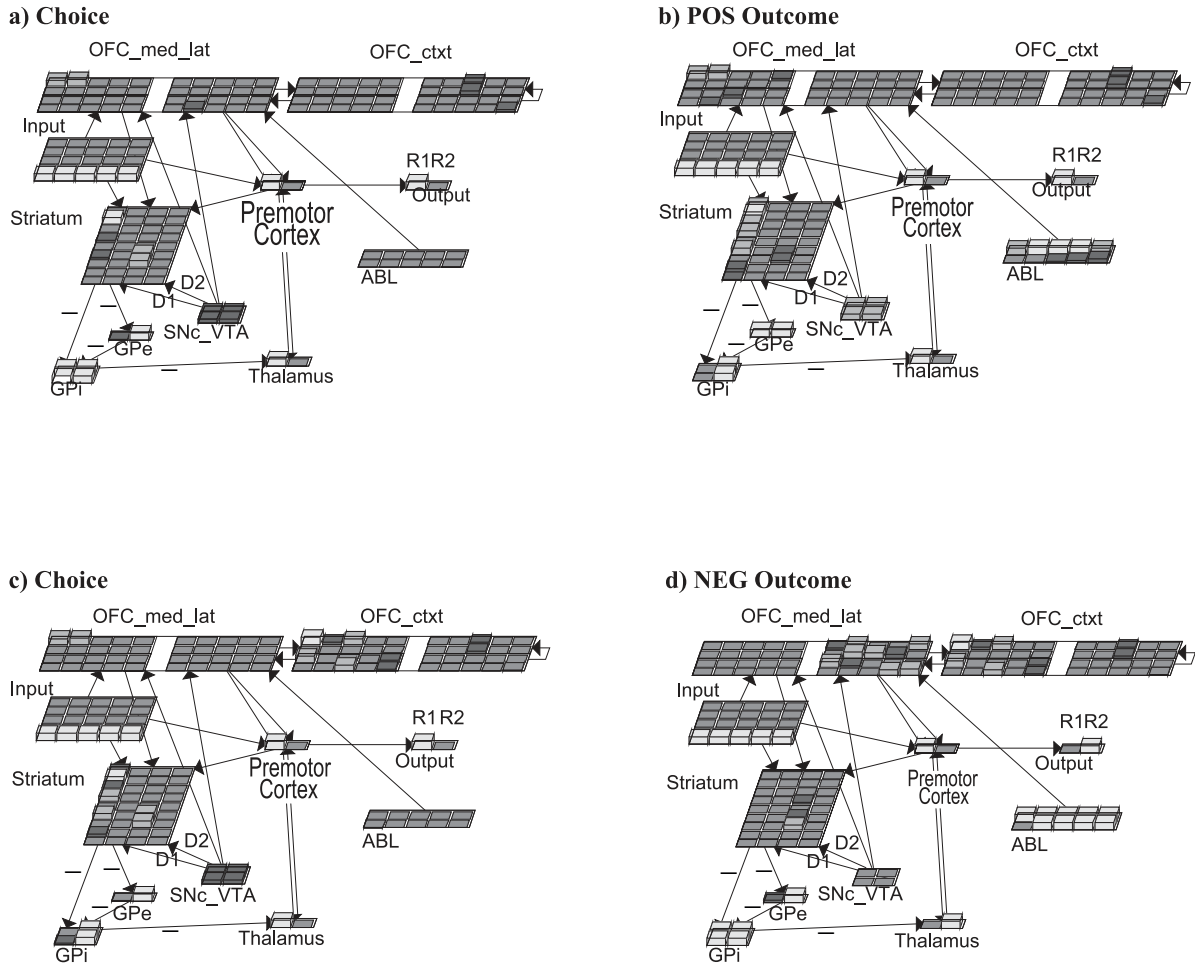
Figure 4.   Gambling task: Positive and negative feedback. a: Deck A is presented to the network, represented by the first row of units in the input. Early in training, the basal ganglia (BG) has not learned to gate either response. The orbitofrontal cortex (OFC) receives direct sensory projections from the input and forms a distributed representation of Deck A in medial and lateral areas. A random greater initial medial (left) than lateral (right) OFC association leads to top-down selection of the play (R1) response. b: The network receives a small gain for its play response, as can be observed by an increase in dopamine firing in the SNc-VTA and moderate magnitude activity in the basolateral nucleus of the amygdala (ABL). The dopamine increase enhances go firing and suppresses no-go firing in the BG, driving learning as in the primitive BG model (Frank, 2005). The OFC receives magnitude information from the ABL and reflects this moderate gain with increased medial and decreased lateral activity, relative to the choice phase. In the subsequent trial, the network will update its OFC_ctxt representations to maintain this reinforcement information in working memory. c: The network is again presented with Deck A in a later trial. Now the BG go representation for play is greater than the no-go representation, facilitating the selection of R1. d: The network receives a large loss, as shown by a large amount of ABL activity and zero dopaminergic firing in SNc-VTA. The resulting lateral OFC representation reflects this large magnitude loss. The network will maintain this loss information in lateral OFC_ctxt in subsequent trials, so that the next time Deck A is presented, OFC_med_lat can reflect this negative experience and can bias responding in the BG and premotor cortex to not play Deck A, even if it often results in small gains. Conversely, Deck B is associated with infrequent large gains, which are the represented in a distributed fashion in medial OFC_ctxt areas together with the lateral representation for Deck A (not shown). POS = positive; NEG = negative; OFC_med_lat = layer that represents current positive/negative outcome expectancy and has top-down biasing effects on response selection processes in both the premotor cortex and striatum; OFC_ctxt = layer that represents contextual information about previously experienced gains and losses by maintaining these in working memory; R1 = Response 1; R2 = Response 2; SNc = substantia nigra pars compacta; VTA = ventral tegmental area; GPe = external segment of the globus pallidus; GPi = internal segment of the globus pallidus.

driven input in prefrontal cortical layer 5 (Seamans, Durstewitz, Christie, Stevens, & Sejnowski, 2001; Wang & O'Donnell, 2001). Moreover, it has also been shown that medial PFC activity increases during the receipt of monetary rewards but decreases when participants do not get the reward that they anticipated (Knutson, Fong, Bennett, Adams, & Hommer, 2003). As elaborated in the Discussion section, the representation of negative reinforcement value in the lateral OFC may also depend on cortical serotonin (5-HT) levels, but we do not consider these here for simplicity.

Top-down modulation of decisions. In addition to representing outcome expectancy, the OFC_med_lat layer has a top-down biasing effect on response selection processes. First, it directly projects to the premotor cortex, which can learn to associate particular medial–lateral representations with output responses (e.g., Wallis & Miller, 2003). Second, OFC reinforcement representations can bias the striatal system to respond go or no-go, allowing the reinforcement representations to take into consideration OFC magnitude representations and not just the simple frequencies of positive/negative reinforcement. Because medial OFC representations co-occur with go activity in the striatum during positive outcomes, Hebbian learning processes result in medial OFC representations tending to bias striatal go responses. Similarly, lateral OFC representations bias striatal no-go responding, preventing the execution of responses that would lead to negative outcomes. There is ample evidence for lateral OFC activation during response inhibition, including no-go trials in go/no-go tasks (Arana et al., 2003; Bokura, Yamaguchi, & Kobayashi, 2001; Horn, Dolan, Elliott, Deakin, & Woodruff, 2003). However, as discussed later, it is likely that some of this no-go behavioral control also depends on interactions with the ACC.

In this manner, magnitude information maintained in medial and lateral OFC working memory representations can exert top-down control over decision-making processes. This control is especially evident when the best decision in a given context is different from what it had been in the past, so that active context information in the OFC can bias the motor system to respond differently from how it had learned by habit. For a schematic of how the different parts of the model contribute to decision making, see Figure 5.

Next, we show that the principles implemented in the model hold in the context of tasks known to depend on OFC integrity: decision making under uncertainty, reversal learning, and devaluation of a conditioned stimulus. The same model parameters were used across all simulations. In addition, the cognitive mechanisms proposed to emerge from the model are able to account for risk-seeking/aversion behavior typically observed in the framing effect.

Results

Decision-Making/Gambling Task

In our gambling task, we tested the model's ability to make decisions based on combined frequency and magnitude of gains and losses. We used a variant of the Iowa Gambling Task (IGT; Bechara, Damasio, Damasio, & Anderson, 1994), in which the option on each trial is to play or pass the next card in a particular deck. This version of the task, based on a previous modification (Peters & Slovic, 2000), has been tested with neurologically intact individuals and gives the same pattern of results as does the original IGT (Claus, 2005).

The model's task was to consider a single input stimulus (*deck*) at a time and to play (R1) or pass (R2). After play responses, the gain–loss outcome was applied to the ABL and DA layers. In all simulations, Deck A was associated with a gain on 70% of trials and a loss on the remaining 30%, and vice versa for Deck B.

In the just frequency condition, the model had to learn to play to Deck A, which is most often associated with a gain, and pass to Deck B, most often associated with a loss. All gains and losses in this condition were set to a magnitude of 1.0. In the magnitude versus frequency condition, the magnitudes of low-frequency outcomes (3.5) were larger than those of high-frequency outcomes (1.0), such that overall expected values were negative for Deck A, despite frequent gains, and positive for Deck B, despite frequent losses.

As expected, the intact model successfully performed across both gambling tasks, always playing the good deck more often than the bad deck (see Figures 6a and 6b). In contrast, when the OFC layer was lesioned, the more primitive BG model successfully performed only in the just frequency condition. In the more challenging task, this model continued to rely largely on frequency information, resulting in greater playing of overall bad than good decks. This result suggests that the BG-DA system is unable to adequately weight the magnitudes of rewards and punishments when making decisions, which are better handled by the more flexible and advanced OFC.

It could be that the BG model's failure to make use of reward magnitude information was due to our use of binary DA signals (i.e., a burst for all positive outcomes and a dip for negative outcomes, regardless of outcome magnitude). This is especially relevant given recent observations that DA neurons fire more with increasing reward value (Tobler et al., 2005), which could potentially provide the BG system with magnitude information. However, there are several reasons to believe that these low-level DA signals are insufficient to precisely encode magnitude information for adaptive behavior, especially when it conflicts with reinforcement frequency information. First, although DA firing increased monotonically with increasing reward value, this effect was not linear: An increase in reward value by 233% elicited a change in DA firing of roughly 130% (see Figure 2 in Tobler et al., 2005). Second, for the BG to adaptively use these DA magnitude values, it would have to be shown that there is a monotonic relationship between the degree of striatal learning and the DA levels. But although the presence of DA is critical for corticostriatal plasticity (Centonze et al., 2001; Kerr & Wickens, 2001), we know of no evidence to support such a monotonic, linear relationship. Because Tobler et al. (2005) recorded DA neurons from intact monkeys, the possibility remains that the increased DA firing for larger reward magnitudes reflects top-down projections from OFC magnitude representations onto DA neurons. This possibility is consistent with the authors' observations that DA neurons encoded not absolute but relative reward values, a property that has also been observed in OFC neurons (Tremblay & Schultz, 1999). Finally, the full range of DA magnitude signals may not be provided by the DA system, especially for negative magnitudes (Bayer & Glimcher, 2005).

Nevertheless, to explicitly address this issue, we also ran a condition in which the magnitude of the DA signal was scaled by the magnitude of outcomes. In these simulations, large magnitude outcomes were associated with maximal positive and negative DA firing (SNc-VTA values of 1.0 and 0.0, respectively, a change of ± 0.5 from tonic 0.5 DA values), whereas small magnitude outcomes were associated with more modest bursts and dips (0.7
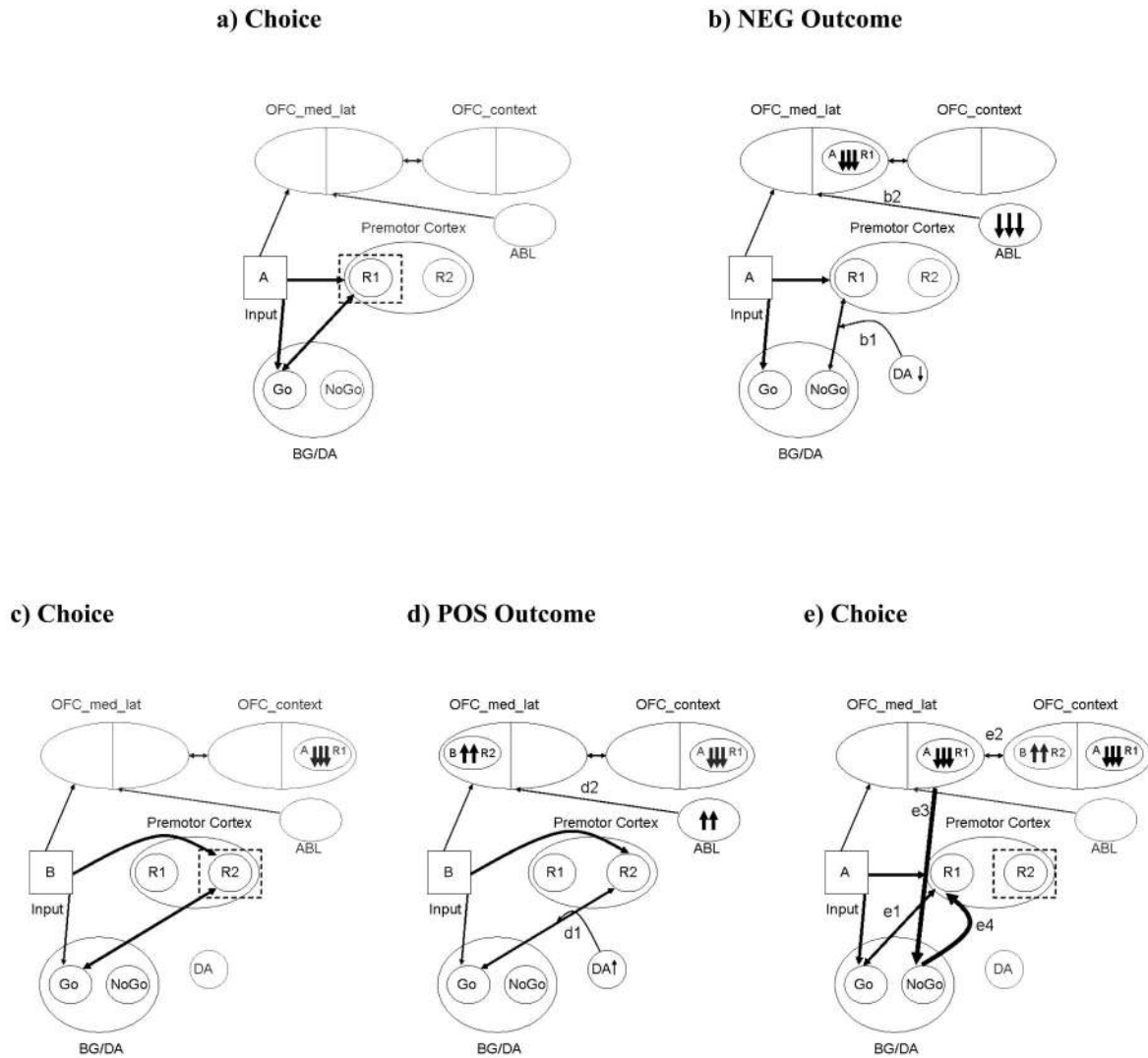
Figure 5. Schematic of model dynamics: Hypothetical sequence of decision trials. a: Stimulus A is presented in the first trial of a new session. Prior ingrained stimulus–response associations lead to Response 1 (R1) selection via strong connection weights from the input and basal ganglia (BG) go cells. b: The outcome is negative (punishment), represented by a dip in dopamine (DA) firing and a negative magnitude representation in the basolateral nucleus of the amygdala (ABL). The DA dip drives no-go learning so that R1 is less likely to be selected in response to Stimulus A on subsequent trials (b1). The lateral OFC encodes the negative magnitude outcome from the ABL together with the stimulus (b2) and also enhances no-go BG firing and learning (not shown). c: Stimulus B is presented, and prior ingrained associations lead to selection of Response 2 (R2). Note that the lateral OFC continues to maintain the negative outcome associated with Stimulus A, but this does not influence responding in this trial because the associated stimulus context (A) is not present in the input. d: The outcome is positive (reward), represented by a burst of DA and increased ABL firing. The DA burst enhances go learning to reinforce the selection of R2 in Stimulus Context B (d1). The medial OFC encodes the reward associated with the stimulus context (d2) and additionally reinforces go firing in the BG (not shown). e: When Stimulus A is presented again, habitual BG–motor responding begins to activate R1 (e1), but then the active lateral OFC representation is retrieved (e2) and exerts a top-down bias on no-go cells in the BG (e3), which, if strong enough, can override (e4) the ingrained BG go-R1 association (e1). The alternative response (R2) becomes more likely to be executed. The Stimulus B–reward association continues to be maintained in the medial OFC and can bias go responding to R2 in subsequent Stimulus B trials. NEG = negative; POS = positive; OFC_med_lat = layer that represents current positive/negative outcome expectancy and has top-down biasing effects on response selection processes in both the premotor cortex and striatum; OFC_context = layer that represents contextual information about previously experienced gains and losses by maintaining these in working memory.
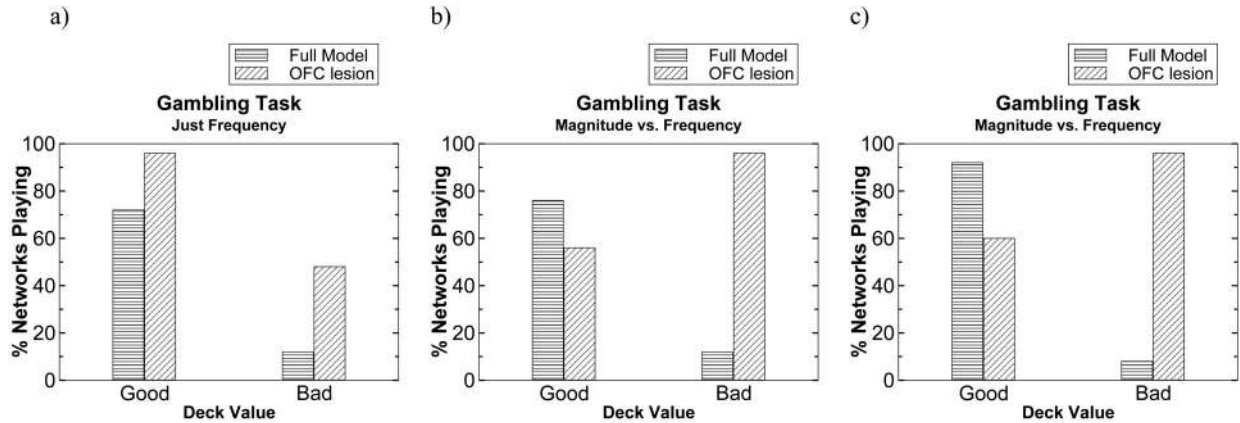
Figure 6. Gambling task results after 140 trials of training. a: In the just frequency condition, both intact and OFC-lesioned models were successful at playing to the good deck (which resulted in a gain 70% of the time) and passing on the bad deck (which resulted in a loss 70% of the time). b: When magnitude information was in opposition to frequency, the full model was nevertheless able to maximize expected value by playing on the infrequent high-gain deck and passing on the infrequent high-loss deck. In contrast, the OFC-lesioned networks continued to respond on the basis of frequency and therefore make maladaptive decisions. c: These results held up even when the dopamine signal was scaled such that high-magnitude gains–losses were associated with larger dopamine changes than were low-magnitude outcomes. Thus, even though the striatum had access to scaled-magnitude dopamine signals, it was still unable to use these signals to override the effects of more frequent dopamine signals in the reverse direction. OFC = orbitofrontal cortex.

and 0.3, ± 0.2 from tonic values). Thus, the BG model now had access to magnitude information and could potentially use this to counteract the frequency information. However, as shown in Figure 6c, scaled DA values produced the same qualitative pattern of results, with the OFC-lesioned model continuing to make maladaptive decisions. This is because infrequent large DA bursts and dips were insufficient to counteract the effects of several instances of smaller DA values in the opposite direction.[4] Of interest, the full model performed modestly better with scaled DA values, as these had concomitant effects on OFC working memory activity for good and bad outcomes.

To explore how these divergent decision-making behaviors evolve across learning, we plotted the difference curves (good–bad deck playing) as a function of trials (see Figure 7). In the just frequency condition, the OFC-lesioned model slowly integrated positive and negative outcomes, eventually outperforming even the full model. In contrast, OFC working memory representations allowed the full model to quickly respond adaptively. However, these same working memory representations caused the model to be more likely to probability match rather than to optimize responding, as is observed in healthy human participants (Estes, 1961). That the OFC-lesioned model is eventually more optimal in this task than is the full model provides an intriguing account for why, after having been trained, rats (with a less developed frontal cortex) are more optimal in their responding than are humans (Hinson & Staddon, 1983). Although people can be induced to respond optimally under some conditions (Shanks, Tunney, & McCarthy, 2002), our model predicts that active maintenance of recent reinforcement values may cause people to probability match more than they would otherwise. Finally, Figure 7 shows that when reinforcement magnitude is in opposition to frequency, the intact model progressively learned the expected value of each deck, whereas the more primitive model increasingly capitalized only on frequency and therefore failed to extract the true expected

value of decisions. Again, scaled DA values reproduce the same pattern of results, albeit with slightly better performance for both intact and lesioned models.

Reversal Learning

As discussed earlier, the OFC is consistently implicated in reversal learning tasks (Rolls, 1999). In our simulations, we presented one of two input stimuli (A or B) to the model, which had to either select this stimulus (R1) or switch to the alternative stimulus (R2). A similar method was implemented to model a two-alternative forced-choice task in previous work (Frank, 2005; Frank, Rudy, & O'Reilly, 2003). Networks had to learn to choose Stimulus A and avoid Stimulus B. After 200 trials, the contingencies were reversed, such that Stimulus A resulted in negative reinforcement and Stimulus B was now associated with positive reinforcement.

Results were consistent with the hypothesized OFC role in reversal learning (see Figure 8). Although both the primitive and

---

[4] Theoretically, a perfect Rescorla–Wagner integrator should be able to use counteracting magnitude–frequency information. However, this assumes that larger DA magnitude values are linearly translated into more learning with a one-to-one relationship. But in the BG model (and likely in the brain), these effects are somewhat nonlinear: Large magnitude DA signals do drive more learning than do small magnitude signals, but the differences in weight changes are not one-to-one. This nonlinearity can arise from a combination of (a) a restricted dynamic range of neural firing or learning (e.g., some go neurons may already be firing close to their maximal rate; large DA bursts could then drive them only to maximal firing, resulting in a loss of information) and (b) the weight contrast enhancement function used in our simulations (O'Reilly & Munakata, 2000), corresponding to nonlinear effects on second messenger systems that drive plasticity.
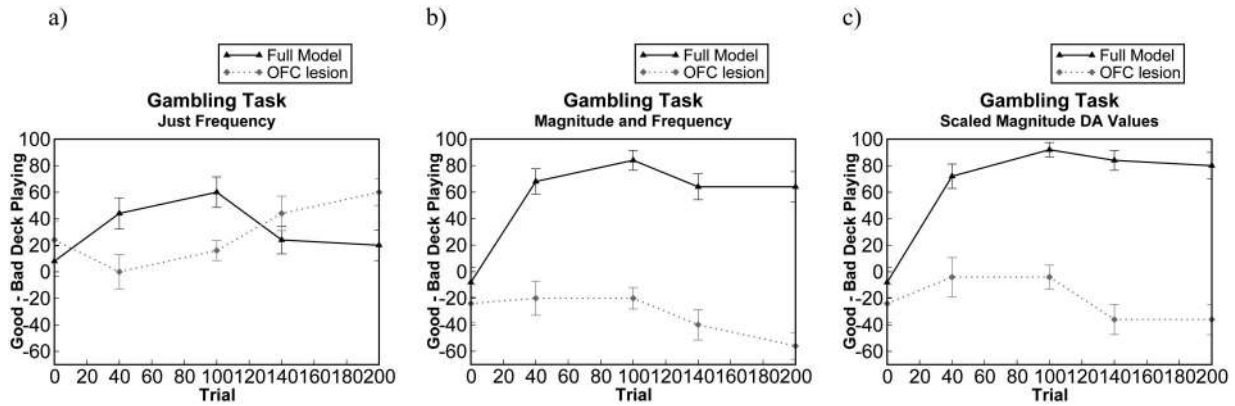
Figure 7. Model performance (good–bad deck playing) in the gambling task across training trials. a: In the just frequency condition, the full model was able to make use of active working memory representations to respond adaptively early in training. Nevertheless, these same working memory representations resulted in increased probability matching as training progressed. In contrast, the OFC-lesioned networks were slower to integrate frequency information over time but responded more optimally later in training. This result is consistent with observations that rats with less frontal cortex are more optimal in their responding than are humans in probabilistic tasks (Hinson & Staddon, 1983). b: The full model takes advantage of magnitude information to respond more to decks with overall higher expected value. In contrast, the OFC-lesioned model progressively incorporates frequency information and therefore makes more maladaptive decisions as training progresses. c: Again, scaling dopamine values by magnitudes of gains and losses produces the same qualitative pattern of results when magnitude information is the more informative decision attribute. Error bars reflect standard error of the mean across 25 networks with different random initial weights. OFC = orbitofrontal cortex.

augmented models were successful at discriminating between the positive and negative stimuli, OFC lesions resulted in selective impairment in the reversal stage, mirroring the results observed behaviorally in humans with OFC damage (Fellows & Farah, 2003). Thus, as in the gambling task, OFC working memory representations are critical for counteracting habitual responses that have been ingrained over multiple experiences. In the gambling task, OFC magnitude representations are necessary to override conflicting frequency information within the same segment. Although magnitude is not a factor in the reversal task, OFC representations nevertheless support adaptive, flexible behavior by enabling rapid shifting of previously ingrained associations.

Devaluation

Another common function attributed to the OFC is representing goal-directed action–outcome contingencies that can override
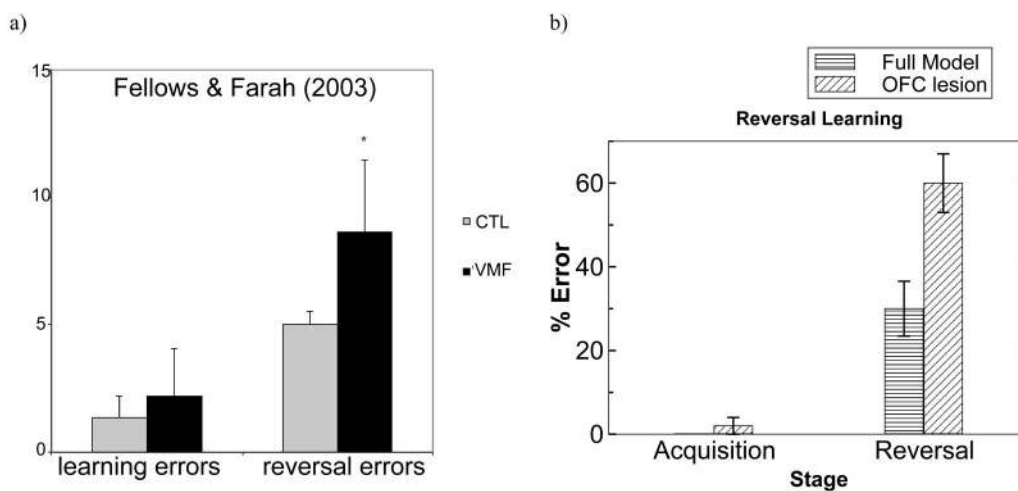
Figure 8. a: Reversal learning impairments in humans with damage to ventromedial and orbitofrontal cortices, showing number of errors made in the learning and reversal phases. Modified from Fellows and Farah (2003) with permission. b: Model reversal learning results. *Acquisition* refers to performance (error percentages) after 200 trials; *reversal* refers to performance after a further 200 reversal trials. Error bars reflect standard errors of the means. OFC = orbitofrontal cortex; CTL = control participants, VMF = patients with ventromedial prefrontal cortex lesions.

stimulus–response habits (Balleine & Dickinson, 1998; Holland & Gallagher, 2004). This is typically studied in the context of a devaluation paradigm. In this procedure, animals must learn to respond to a conditioned stimulus to obtain reward. Once this learning is achieved, the reinforcer is subsequently devalued (either by means of satiation or by making the animal sick with lithium chloride). Instrumental performance is typically goal-directed early in training and is therefore sensitive to outcome devaluation. However, extensive training results in stronger influences of the stimulus–response ingrained habit, and behavior is no longer sensitive to devaluation (i.e., animals will continue to press a lever that leads to the devalued outcome).

Several lines of evidence suggest that the OFC supports goal-directed behavior and is critical for devaluation, whereas the striatum supports habitual responses. First, rats and monkeys with OFC lesions are not sensitive to devaluation (Gallagher, McMahan, & Schoenbaum, 1999; Izquierdo, Suda, & Murray, 2004; Pickens, Saddoris, Gallagher, & Holland, 2005). In contrast, rats with dorsolateral striatal lesions are more sensitive than control rats (Yin, Knowlton, & Balleine, 2004). Second, there is evidence that the OFC learns action–outcome representations via input from the amygdala (Pickens et al., 2003). Third, neuroimaging studies show changes in OFC activity following devaluation in humans (Gottfried, O'Doherty, & Dolan, 2003; Small, Zatorre, Dagher, Evans, & Jones-Gotman, 2001). In one of these studies, medial OFC activity was observed when participants began eating chocolate, but as a sated state was reached, the lateral OFC became more active (Small et al., 2001).

To explore whether the same principles that explain BG-OFC function in gambling and reversal learning tasks could also be applied to devaluation, we trained the models with a devaluation procedure. All models were presented with 100 trials of an input stimulus followed by positive reinforcement for selecting R1. Subsequently, this stimulus was devalued by pairing it with five trials of negative ABL magnitude signals and DA dips, in the absence of motor responding. Finally, the models were tested to see whether they would continue to respond with R1 in the presence of the stimulus.

Results were consistent with available data (see Figure 9). After 100 trials of training, all models responded with 100% accuracy to the conditioned stimulus. Following devaluation, intact networks were less likely to respond to the stimulus, showing sensitivity to action–outcome contingencies. This sensitivity was abolished in OFC-lesioned networks and in intact networks with more extended training (300 trials) prior to devaluation, both showing habitual responding to the stimulus. Finally, networks with striatal lesions were more sensitive to devaluation, consistent with results observed empirically (Yin et al., 2004). Thus, our model formally shows that the OFC can support flexible action–outcome behavior, whereas the striatum is important for ingraining stimulus–response habits into motor cortical areas. We return to the issue of how these systems compete for behavior in the Discussion section.

Risk Aversion and Risk Seeking

Although the above simulations successfully simulated the effects of localized brain lesions on learning and decision making, it may not be readily apparent how this model could account for traditional decision-making patterns observed in healthy individuals. Here we apply the model to provide a neurocomputational
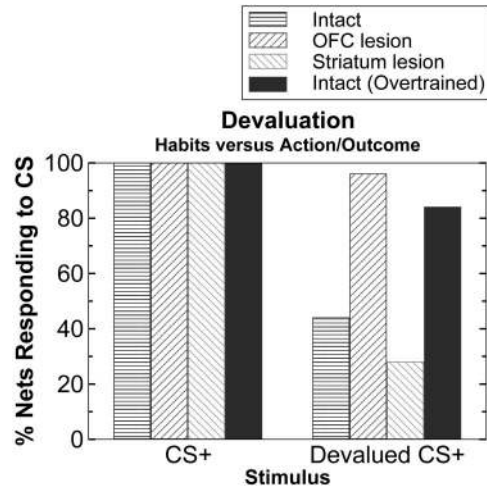
Figure 9. Orbitofrontal cortex (OFC) and basal ganglia contributions to representing stimulus–response habits versus higher order action–outcome contingencies. After 100 trials of positive reinforcement to a stimulus, all networks (Nets) successfully responded to this stimulus in 100% of cases. After devaluation, only 44% of intact networks responded to the stimulus, showing devaluation. Networks with OFC lesions continued to respond on the basis of habitual associations. In contrast, networks with striatal lesions showed enhanced sensitivity to devaluation, with only 28% responding to the stimulus. Finally, after more extended training, even intact networks responded habitually and were insensitive to devaluation. This pattern of results is consistent with those observed empirically in animals with OFC and striatal damage (Gallagher et al., 1999; Izquierdo et al., 2004; Yin et al., 2004). CS = conditioned stimulus.

account of classical risk-seeking/aversion behaviors, as described by the framing effect (Tversky & Kahneman, 1981; see Figure 10a). Decision makers are asked to choose among two options, for which the first has a certain outcome but the second has a probabilistic outcome. The expected value of each option is typically the same. But depending on how the decision is framed (in terms of gains or losses), different choice patterns are observed. If the outcomes are potential gains, the certain option is preferred over the probabilistic option (e.g., people choose a certain gain of $100 instead of a 50% chance of $200); this phenomenon is referred to as risk aversion (Kahneman & Tversky, 1979). But when the decision options are framed in terms of losses, participants generally prefer the risky option (e.g., they choose a 50% chance of losing $200 over a certain loss of $100); that is, they show risk-seeking behavior.

These differences in risk seeking for losses and gains have been accounted for by *prospect theory* (Kahneman & Tversky, 1979). The central tenets of prospect theory with respect to magnitudes of gains and losses are that (a) losses have a larger effect on underlying value than the equivalent magnitude gains and that (b) people evaluate gains or losses from some neutral or reference point (in this case, the certain outcome). For example, when compared with a certain outcome of $100, the probabilistic option of gaining $200 or $0 is rerepresented as a relative gain of $100 and a relative loss of $100 rather than only a potential gain. That prospect theory posits relative reward values in decision-making computations lends itself naturally to an OFC mechanism: As discussed previously, OFC neurons are known to represent relative, rather than
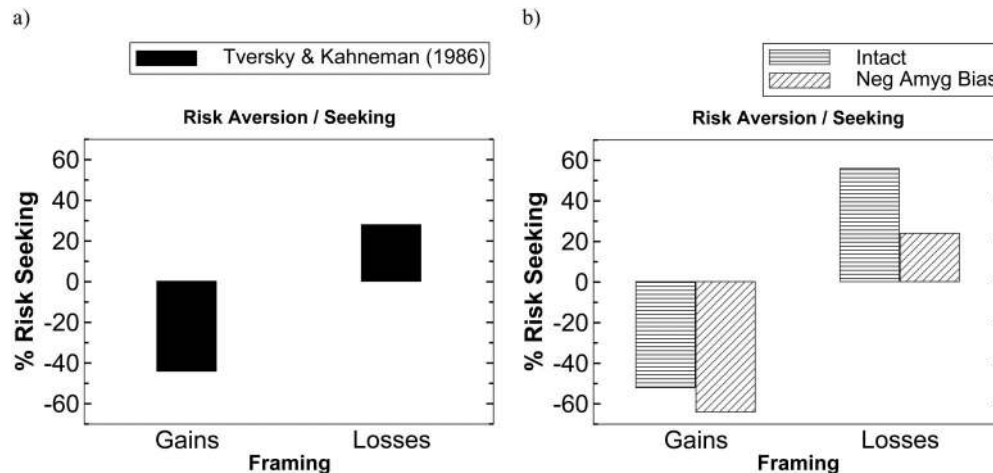
Figure 10. a: Classical framing effects show that humans are risk aversive for gains (e.g., they tend to choose a certain $100 win over a 50% chance to gain $200 and a 50% chance to gain nothing) but are risk seeking for losses (they avoid a certain loss of $100 in favor of a 50% chance to lose nothing and a 50% chance to lose $200). Data plotted from numerical results described in Tversky and Kahneman (1986). b: Intact networks replicated this pattern of results: They were risk averse for gains and risk seeking for losses. Plots show percentages of certain option choices subtracted from percentages of risk option choices for gains and losses, respectively. When networks were made more sensitive to losses than to gains by increasing relative amygdala magnitude representations for negative compared with positive outcomes (Neg Amyg Bias), they showed the asymmetrical effect observed in the literature, in which participants are more risk aversive for gains than they are risk seeking for losses.

absolute, reward preferences (Tremblay & Schultz, 1999). As described by Plous (1993),

> If the reference point is defined such that an outcome is viewed as a gain, then the resulting value function will be concave and decision makers will tend to be risk averse. On the other hand, if the reference point is defined such that an outcome is viewed as a loss, then the value function will be convex and decision makers will be risk seeking. (p. 97)

When simulated in our model, these same decision-making framing effects were observed (see Figure 10b). For gains, we trained networks with two decks, one of which was associated with a positive outcome on 100% of trials (certain gain, i.e., the reference point) and another of which was associated with a relative gain on 50% of trials and a relative loss on the remainder of trials (these relative gains–losses were of equivalent magnitudes to the certain gain, as in the example described above). Options that led to certain gains became associated with medial OFC representations, causing the networks to be biased toward selecting that option. When faced with the alternative, more risky option, networks represented the relative OFC outcomes as partially medial and partially lateral, owing to the ambiguous chances of winning or losing relative to the sure outcome. The net result was that the networks were risk aversive; that is, they were more likely to choose options leading to the certain gain.

Similar reasoning explains why the model was actually risk seeking for losses. These simulations were identical to those above, except that the deck with a certain outcome resulted in a loss (instead of a gain) in 100% of trials. This option became associated with a lateral OFC representation, causing networks to be biased toward avoiding that option. The alternative, more risky option was again associated with a more ambiguous medial and

lateral OFC representation. The net result was that the model avoided the certain loss so that it had some potential of losing nothing, even though it could potentially incur a larger loss. Note that this explanation implies that the risk-seeking behavior observed in humans under the prospect of losses stems from their tendency to avoid certain losses rather than a propensity for risky choices per se. This is consistent with observations that participants are slower to make choices in negative compared with positive framings (Gonzalez, Danda, Koshino, & Just, 2005; Payne, Bettman, & Johnson, 1993), because it should take longer to avoid the certain (negative) option and then consider the alternative option than to just directly choose the certain (positive) option.

Finally, we explored whether our model could account for more nuanced effects showing that people are relatively more risk averse for gains than they are risk seeking for losses (e.g., Gonzalez et al., 2005; Tversky & Kahneman, 1986; Figure 10a). We hypothesized that this asymmetrical effect could result from a relatively greater sensitivity to negative compared with positive affective values (Tversky & Kahneman, 1991), which could in turn stem from a greater representation of negative than of positive information in the amygdala. Although the amygdala is now known to be involved in reward processing (Baxter & Murray, 2002), its classic and well-established role in negative affect (LeDoux, 2000) suggests that negative value might be better represented than is positive value. This bias could also be a result of the more dense amygdalar projections to the lateral than to the medial OFC (Carmichael & Price, 1995; Cavada et al., 2000). To explore this account, we reran the simulations reported above but with all negative outcomes represented with two times greater amygdalar magnitudes than those of positive outcomes.

Simulation results under these conditions reproduced the asymmetrical effects: Networks showed greater risk aversion for gains than they did risk seeking for losses (see Figure 10b). This was because greater negative magnitude amygdala representations resulted in a lateral OFC bias when processing risky options. (Recall that under standard model parameters, these risky options were associated with offsetting medial and lateral values.) This lateral bias caused enhanced risk aversion for gains, because networks were now more likely to avoid the potential (relative) loss associated with the risky option, and this effect was additive with the networks' continued tendency to be biased toward the certain option (which still had a purely medial representation). Further, the same amygdala manipulation diminished the risk-seeking effect for losses: The lateral OFC bias caused networks to now have a greater tendency to avoid the potential of additional losses associated with the risky choice. Nevertheless, because the certain loss was associated with a purely lateral OFC representation, networks continued to reliably avoid this option and continued to display relative risk-seeking behavior.

In summary, our OFC model displayed classical framing effects in that it was risk averse for gains and risk seeking for losses. These effects emerged naturally from our framework because medial OFC representations for certain gains biased networks to choose these options, whereas lateral OFC representations for certain losses biased networks to avoid those options. The risky choices were associated with offsetting medial and lateral representations, as their potential outcomes were encoded relative to the certain gains, as posited by prospect theory (Tversky & Kahneman, 1991). Finally, our account for asymmetrical effects of risk aversion and seeking depends on greater (amygdala) magnitude representations for negative than for positive outcomes. This account is consistent with observations that risk-seeking/aversion behaviors become symmetrical when the risky options are associated with relatively smaller losses than gains (Tversky & Kahneman, 1986).

Discussion

This work represents a unique integration of the contributions of the BG, amygdala, and ventromedial and orbitofrontal cortices to decision making. Through explicit computational simulations, we established a framework for exploring complementary roles of these cortical and subcortical structures. We showed that the more primitive BG-DA system is sufficient for (relatively slow) learning to make choices based on their frequencies of positive versus negative reinforcement. However, when decision-making tasks required more complex strategies, active representations in the OFC are necessary to provide top-down, goal-directed biasing on the decision outputs. In so doing, our model accounts for decision-making deficits observed in patients with OFC damage, in both gambling and reversal learning tasks. Specifically, our model suggests that the core decision-making deficit in these patients is in assigning reinforcement value to decisions based on the magnitude and recent temporal context of expected outcomes. OFC lesions may lead to decision-making deficits when this information has to override the long-term probability of positive/negative reinforcement computed by the BG-DA system. This same model accounts for reversal learning impairments in OFC patients, as very recent temporal context must be used to override prepotent tendencies. Further, the model also provides insight into the differential roles of the OFC and BG in representing goal-directed action–outcome contingencies versus stimulus–response procedural habits. Finally, we showed that the intact network exhibits standard risk-aversion and risk-seeking decision-making patterns that are classically observed in human participants.

The basis for the BG-DA portion of our model was drawn from various pieces of neurobiological data, as reviewed in Frank (2005). In brief, two BG pathways are thought to independently facilitate or suppress cortical motor commands (Mink, 1996). More specifically, two main projection pathways from the striatum go through different BG output structures on the way to the thalamus and up to the cortex (see Figure 1a). Activity in the direct pathway sends a go signal to facilitate the execution of a response considered in the cortex, whereas activity in the indirect pathway sends a no-go signal to suppress competing responses. DA modulates the relative balance of these pathways by exciting go cells while inhibiting no-go cells via D_1 and D_2 receptors, respectively (Gerfen, 2000; Hernandez-Lopez et al., 1997, 2000). This effect is dynamic, such that transient DA increases during positive reinforcement (Schultz, 2002) preferentially activate go cells while suppressing no-go cells, and vice versa for DA decreases during negative reinforcement (Brown et al., 2004; Frank, 2005). This change in activity modifies synaptic plasticity (Centonze et al., 2001), driving learning go for rewarding choices and no-go for maladaptive decisions.

The basis for our implemented OFC mechanisms was similarly motivated by biological and functional data. As reviewed above, medial and lateral OFC areas differentially respond to positive and negative valences (Gottfried et al., 2002; O'Doherty et al., 2001; Rolls & Kringelbach, 2003). Further, OFC activity is correlated with the magnitude of reinforcement outcome, which likely reflects afferent activity from the ABL (Holland & Gallagher, 2004). The OFC then maintains this reinforcement information over time in working memory via persistent neural firing (K. Hikosaka & Watanabe, 2000), which is enhanced by DA (Seamans & Yang, 2004). This active maintenance has top-down biasing effects on behavior (Miller & Cohen, 2001), via efferent projections to the striatum and motor cortical areas (Alexander et al., 1990; Haber et al., 1995).

Emergent properties of our combined BG-OFC model are also consistent with theoretical perspectives on the functional roles of these areas in learning and decision making. First, various accounts have implicated the BG in the development of stimulus–response habits (O. Hikosaka, 1994; Jog et al., 1999; Packard & Knowlton, 2002). Our model suggests that the learning of these habits is accomplished via dopaminergic modulation within go/no-go BG pathways but that these are only later transferred or ingrained in the cortex (Frank, 2005). This prediction was recently supported by observations of learning-related activity occurring earlier in the striatum than in the frontal cortex (Delgado et al., 2005; Pasupathy & Miller, 2005). Further support for a BG-mediated habit learning system comes from observations that striatal lesions lead to less habitual responding and enhanced sensitivity to reinforcer devaluation (Yin et al., 2004), a result that was also borne out in our model. Second, our model also provides a mechanistic account for how the OFC may be involved in representing flexible action–outcome contingencies (Gallagher et al., 1999; Izquierdo et al., 2004), by maintaining recent action–outcome expectancies in working memory and providing top-down influences on the decision outputs. This is a more specific

instantiation of the widely held notion that the PFC is involved in flexible behavior, planning, and decision making (e.g., Fuster, 1989; Miller & Cohen, 2001; Norman & Shallice, 1986).

This conceptualization is also consistent with animal experiments showing that OFC integrity is necessary for learning to make decisions that lead to larger, delayed rewards instead of smaller, immediate rewards (Mobini et al., 2002). The implication of our model is that learning about delayed rewards depends on working memory for action–outcome contingencies and requires suppression of responses that would lead to immediate rewards. Recent neuroimaging results in humans support this account, showing medial OFC activity during selection of an immediate reward (go) and lateral OFC activity when participants suppressed (no-go) this choice in favor of a later delayed reward (McClure, Laibson, Loewenstein, & Cohen, 2004). Activation of the medial and lateral OFC was further observed in an imaging study when participants chose options having a low probability of resulting in large rewards–punishments versus those having a high probability of smaller magnitude outcomes (Rogers et al., 1999). Finally, our suggestion that the OFC is critical for representing reinforcement magnitude in the face of conflicting frequency information is supported by observations that OFC lesions lead to enhanced choice of smaller, more certain rewards instead of less probable but larger rewards (Mobini et al., 2002).

Functional Level Description

Although we have simulated a range of decision-making and reinforcement-learning phenomena with a unified model, one might argue that this model is more complex than it ought to be for capturing these psychological phenomena. We believe that our explicit formulations about how each brain region contributes to decision making has potential to inform the cognitive and clinical neuroscience literature and that this would be more difficult using simpler, more abstract mathematical models. Nevertheless, a functional level description of our model may be useful to evaluate the overall informational processing computations that should hold irrespective of our particular implementation.

At the functional level, we propose two systems that can either work in tandem or compete against each other in biasing the response selection process. The first, and arguably more primitive, system is one that learns to bias responses on the basis of probabilistic information pertaining to rewarding and punishing outcomes. This process emerges from the interaction of the BG-DA system with the premotor cortex. When a stimulus is initially encountered, a prediction of its reward or punishment value based on random initial weights is computed and subsequently biases a response. Learning about the consequences of that particular response occurs during positive or negative reinforcement, which support go or no-go learning about the response just executed. If the outcome is positive, a positive learning signal is produced, which has the effect of strengthening associations to responses that led to the positive outcome. In contrast, negative learning signals increase weights that will subsequently result in suppression of that response, by increasing connection weights to no-go representations. Over the course of time, the probabilistic nature of the outcome structure will result in the model strengthening go representations for stimulus–response conjunctions that are likely to be rewarding and strengthening no-go representations for those that

are not (or that are punishing). Thus, this system can be thought of as a habit learning system.

The other, more advanced system computes reward or punishment magnitude, which is then held online in working memory to bias response selection processes.[5] This magnitude system relies on the interaction of the amygdala, OFC, and BG. Our model suggests that intensity (magnitude) of an outcome is represented by increased neuronal firing in the amygdala, with two different subpopulations of neurons firing for positive and negative outcomes. In addition, information about the valence of the outcome comes from a prediction error, which is computed within the first system. In other words, if the outcome is more positive than expected, the learning signal will be positively valenced, and it will be negative if the outcome is more negative than expected. We also propose that positive and negative outcomes are held separately in working memory and have differential effects on biasing response selection processes, with positive magnitudes biasing go responses and negative magnitudes biasing no-go responses. Because both systems work to bias response selection processes as well as each other, we suggest that overall perceived expected value of a stimulus is an emergent computation that results from the two systems interacting. Thus, behavior that is considered irrational (inconsistent with expected values) may result from the relative biases of the two systems on response selection processes.

Our model also posits an explicit mechanism for how these two systems interact and compete in decision making (for similar arguments on emergent competitive dynamics between the BG and hippocampus in memory, see Atallah, Frank, & O'Reilly, 2004; Frank, O'Reilly, & Curran, in press). Specifically, during initial stages of a task, competing responses are simultaneously activated (considered) in the premotor cortex, and the BG system learns to facilitate adaptive responses and to suppress those that are maladaptive through trial and error. During this stage, active orbitofrontal representations encoding recent reinforcement magnitudes of alternative decisions can exert a considerable amount of control over behavior by directly biasing premotor responding (and potentially also by influencing the BG). If an outcome is subsequently devalued, this change is represented in the OFC and the animal modifies its behavior accordingly, allowing it to override the previously learned response (see Figure 4e for a depiction of this situation). But as learning progresses and the habitual response is repeatedly chosen, the premotor cortex itself becomes biased to preferentially activate (i.e., generate) this response, without requiring as much facilitation from the BG. This is consistent with observations that striatal firing during well-learned responses occurs after that in premotor cortical areas (e.g., Alexander & Crutcher, 1990b; Crutcher & Alexander, 1990; see also Mink, 1996). During this latter habitual responding stage, OFC representations exert less control over behavior, because these representations have to compete with direct premotor activation, and the additional neural processing associated with the OFC circuit is temporally delayed, making it more difficult to override ingrained behaviors. Thus, even intact animals exhibit habitual responding

---

[5] Working memory in the context of the model does not necessarily imply information that is consciously accessible. Rather, the stimulus–outcome contingencies are represented as persistent neural firing that decays gradually and can exert top-down influences even in the absence of awareness.

after extended training. Nevertheless, primates and humans with more developed and stronger frontal–cortical representations continue to be sensitive to action–outcome contingencies.

Relation to Other OFC Models

Other computational models of OFC function have focused on the role of the OFC in reversal learning (Deco & Rolls, 2005), decision making (Wagar & Thagard, 2004), and attentional set shifting (O'Reilly et al., 2002). All of these models explore specialized mechanisms within the OFC that support such behavior. However, none of these models provide a unified account that reconciles disparate theories of OFC function across multiple tasks. Moreover, our model is the first to address the division of labor between the BG-DA system and the OFC–amygdalar system and how their emergent interactions may influence decision making. This integrated model successfully accounted for empirical data across decision-making, reversal learning, and devaluation paradigms, without changing any model parameters. We address specific points of overlap and contrast among the various models next.

In the Deco and Rolls (2005) model, OFC neurons maintain stimulus–reinforcement rules in working memory, similar to their general role in our model. Their model diverges, however, in the mechanisms used to support reversal learning. Specifically, error detection neurons send a global inhibition signal and cause simulated OFC neurons to reverse responding. Thus, if Stimulus 1 is initially associated with reward and Stimulus 2 is associated with punishment, a subsequent error signal causes OFC neurons to reverse their selectivity, such that Stimulus 2 is then automatically associated with reward. Thus, although their model implements a potential mechanism by which OFC neurons reverse responding following errors, it does not address the difficult problem of how OFC representations develop or learn in the first place. Further, their model's mechanisms are limited to the simple case of two stimuli, for which one is rewarded and one is not; in the current state, it is unclear how it would scale to accommodate multiple stimulus–reinforcement mappings. In contrast, our model develops distributed and learned OFC representations of reinforcement value, which do not fundamentally restrict the number of possible mappings.

In the Wagar and Thagard (2004) model, the functions of ventromedial–orbitofrontal and amygdala neurons are simulated in a decision-making task. These areas are trained with good–bad affective states that ultimately influence processing in the NAc. If the two affective states are inconsistent with each other, the ventromedial–prefrontal cortex (VMPFC) state trumps that of the amygdala and has more of an influence on the NAc. Thus, VMPFC lesions lead to an oversensitivity to immediate rewards, which are represented in the intact amygdala. In a broad sense, this is similar to our model in that OFC and subcortical areas have complementary and competitive effects on output responding. However, in our model, OFC representations depend on, rather than compete with, input from the amygdala (Holland & Gallagher, 2004); the competition for output responding emerges via interactions with the dorsal striatum and its effects on response selection processes via BG output structures, the thalamus, and the premotor cortex. Further, their model does not learn its VMPFC representations, which instead are applied directly by the modeler. Thus, the VMPFC does not have to maintain information in working mem-

ory, and its effects on the NAc are somewhat of a foregone conclusion. In contrast, complementary roles of the BG and OFC in our model emerge as a result of training—our simulated OFC learns positive/negative associations via trial and error, maintains them in working memory, and influences both the BG and motor cortical areas. Nevertheless, our model does not include contributions of the NAc, which are likely to be critical. In particular, although amygdala activity directly influences the OFC in our simplified model, we think that the NAc and/or ventral striatum may actually be involved in gating of this reinforcement magnitude information to be maintained in the OFC (cf. Frank et al., 2001).

Our model is both consistent with and differs from that of O'Reilly et al. (2002) in several important ways. Like our model, their simulated PFC-OFC has top-down biasing effects on behavior (Miller & Cohen, 2001). But in their particular implementation, the currency of OFC information is in representing detailed stimulus features and not their reward values. Our model is therefore more grounded by the experimental literature, which consistently implicates OFC representation of reward values across both single-cell recordings in primates and imaging studies in humans (Kringelbach & Rolls, 2004; O'Doherty et al., 2001; Tremblay & Schultz, 1999; Wallis & Miller, 2003). It is also unclear how their model would account for decision-making deficits associated with OFC damage. Finally, whereas their model explores divergent roles of the OFC versus the dorsolateral PFC in reversal and set shifting, our model focuses on medial and lateral OFC interactions with the BG, which was omitted in O'Reilly et al. (2002).

In summary, our model represents a unique integration of neural and behavioral data that has not been previously explored and accounts for several pieces of existing data. Next, we discuss particular implications of this framework for decision-making deficits in clinical populations, before elaborating on further novel predictions.

Implications for Clinical Populations

Our model suggests that decision-making deficits in OFC patients stem from degraded working memory representations for the magnitudes of recent reinforcement outcomes. By this account, patients should be able to use their intact BG-DA system to successfully integrate positive/negative experiences over multiple experiences. Patient deficits should be more marked when the magnitude of reinforcement outcomes plays more of a role in determining the expected value of a decision than does the frequency of positive versus negative outcomes. In particular, although the magnitude of infrequent gains and losses may be available to the BG-DA system in terms of larger DA spikes and dips (Tobler et al., 2005), these may not be sufficient to counteract the learning effects of several instances of smaller magnitude outcomes. Thus, patients with OFC damage are predicted to fail at the IGT because of a failure to integrate both magnitude and frequencies of gains versus losses. Because the IGT was not designed to test this possibility (i.e., outcomes always include a gain, and the largest magnitude outcomes are in the bad decks), we are developing modified versions of the task to do so.

Our model may also shed light on decision-making deficits in other clinical populations. For example, drug abusers perform poorly on decision-making tasks, showing inability to maximize long-term rewards (Bolla et al., 2003; Grant, Contoreggi, & Lon-

don, 2000). These deficits were initially hypothesized to result from either hypersensitivity to immediate rewards or generally insensitivity about the future (Bechara, Dolan, & Hindes, 2002). Mathematical modeling of choice behavior on the IGT showed that abusers appear to be insensitive to punishments (Busemeyer & Stout, 2002; Stout, Busemeyer, Lin, Grant, & Bonson, 2004); the model also included parameters for memory of past consequences and response thresholds. Our model is broadly consistent with this characterization but specifies the neural bases for these effects: Reward and punishment sensitivity is determined by DA modulation and medial–lateral OFC representations, memory for recent context is implemented in OFC_ctxt, and response threshold is dictated by the balance of go/no-go activity in the BG. Thus, our model would suggest that an insensitivity to punishment in abusers may be associated with reduced lateral OFC activity, consistent with recent observations in stimulant abusers showing reduced OFC metabolism (Volkow et al., 2001). Although drug abuse could feasibly be an effect rather than a cause of impaired OFC function, indirect evidence from animals shows that cocaine administration can cause insensitivity to action–outcome contingencies in a devaluation paradigm (Schoenbaum & Setlow, 2005), which also depends on the lateral OFC in our model. Finally, the implication of the lateral OFC in delayed gratification (McClure et al., 2004) may explain why drug abusers discount the future at almost twice the rate of non-drug abusing individuals (Bickel & Marsch, 2001).

In addition to predicting deficits in drug abusing participants, our model may help to explain normal individual differences in decision-making abilities. For example, our simulations of classical framing effects suggest that the medial OFC should be preferentially involved in risk-averse behavior to achieve sure gains, whereas the lateral OFC should be activated during risk seeking to avoid certain losses.

Further Model Predictions: Patient and Neural Data

Our model makes several predictions about the nature of decision-making deficits that would be observed in PD and OFC patients, with regard to frequency and magnitudes of punishments and rewards. We also outline some neurophysiological predictions of the model.

PD patients. Our models predict that PD patients should have particular trouble learning to choose advantageously when options differ only in the frequency of rewarding and punishing outcomes, due to reduced dynamic range of BG-DA signals (Frank, 2005). In simple reinforcement learning paradigms, patients are not always globally impaired (Frank et al., 2004), but the BG model nevertheless predicted the differential pattern of learning biases depending on the patient's medication status. Specifically, patients off medication were better at no-go learning from negative reinforcement, whereas medicated patients were more sensitive to positive outcomes but were impaired at learning from negative outcomes. The overall lack of impairment in PD patients compared with control participants could potentially be explained by intact patient maintenance of reinforcement information in the OFC, especially because this frontal area interacts with ventral striatal areas that are relatively spared in PD. Thus, our model predicts overall PD impairments in decision-making tasks with multiple reinforcement magnitude outcomes, in which the frequency of rewards and punishments is more informative than their individual magnitudes.

Further, DA medication may block DA dips and reduce lateral OFC representations, which in the model would prevent the encoding of loss magnitude information while sparing or enhancing gain representations—potentially explaining the documented effects of medication on inducing pathological gambling in PD (Dodd et al, 2005). Patients should also have general impairments in more complex decision-making tasks that involve high working memory load (i.e., in a gambling task with multiple decks and corresponding reinforcement outcomes), as reduced striatal DA in PD is associated with prefrontal deficits similar to those observed in patients with frontal damage (Muller, Wachter, Barthel, Reuter, & von Cramon, 2000; Remy, Jackson, & Ribeiro, 2000).

OFC patients. In contrast to PD patients, OFC patients are predicted to show greater difficulty when choosing between options that differ in magnitudes of rewards. In particular, medial lesions to the OFC are predicted to result in a difficulty of learning a go response to options in which the reward is relatively small but frequent, whereas lateral lesions to the OFC should result in a difficulty learning a no-go response to options that lead to long-term losses, particularly if these decisions will lead to short-term gains.

Our model also makes the counterintuitive prediction that OFC patients should eventually respond more optimally than do control participants in a pure frequency-based reinforcement learning paradigm. This prediction stems from our OFC model's tendency to probability match as a result of working memory representations for gains and losses in the infrequent trials having an effect on subsequent performance. In contrast, the BG system should eventually respond more optimally on the basis of the highest probability of positive reinforcement, despite being slower to acquire this knowledge. This type of prediction, for which OFC patients should perform better than neurologically intact control participants, has been supported in a recent study in which OFC patients performed more optimally than did control participants on an investment task in which the rational choice was more risky (Shiv, Loewenstein, & Bechara, 2005). Nevertheless, this prediction depends on the time constant under which OFC integrates previous reinforcement experiences, such that short time constants (fast decay of working memory) would predict probability matching behavior but longer time constants would lead to more integration.

Neurophysiological predictions. At the neural level, our model predicts that the medial OFC has top-down biasing effects on striatal go signals, whereas the lateral OFC activity should bias no-go signals. Thus, if multiple-cell recording procedures were used during a reinforcement learning paradigm, we would expect medial OFC activity to be correlated with striatal activity in the direct pathway (striatonigral neurons). Conversely, lateral OFC activity would be expected to correlate with activity in the indirect pathway (striatopallidal neurons). Although it is difficult to discern direct from indirect pathway neurons with extracellular recordings, these can be inferred indirectly (Onn, West, & Grace, 2000).

Model Limitations and Further Neurobiological Considerations

Although our model accounts for many empirical findings, we are not suggesting that it is a comprehensive model that will explain all decision-making findings but rather a framework with which to explain interactions between decision making and reinforcement learning. As mentioned previously, we have focused

only on a distinct subset of brain regions that have been implicated in the literature. Below, we elaborate on some of these shortcomings and how future work might incorporate these functions.

Insula. The insula–somatosensory cortex has been proposed to be important for the conscious experience of emotions (Damasio, 1994) and in linking cognition with visceral states (Mesulam, 2000). Several neuroimaging studies have found increased insula activation during decision-making and reversal learning tasks (Ernst et al., 2002; Paulus, Rogalsky, Simmons, Feinstein, & Stein, 2003; Remijnse, Nielen, Uylings, & Veltman, 2005). Because we were not attempting to model the conscious experience of emotions, this portion of the decision-making circuit was omitted. However, future computational models should attempt to incorporate the development of representations of body states that accompany emotions and that may influence decision making (e.g., Bechara & Damasio, 2005).

ACC. The present model also does not explicitly simulate the contributions of the ACC, which is critically involved in negative reinforcement learning (Frank et al., 2005; Holroyd & Coles, 2002) and in monitoring conflict among multiple decision options (Yeung, Botvinick, & Cohen, 2004). Researchers have previously argued that the ACC may support aversive no-go learning by interacting with the BG (Frank et al., 2005). Further, external activation of ACC cells has been shown to produce an aversive teaching signal (Johansen & Fields, 2004), and ACC activity predicts behavioral switching from an unsuccessful response to a more adaptive response; this switch does not occur following ACC inactivation (Shima & Tanji, 1998). Future research is required to disentangle the division of labor between the ACC and lateral OFC in avoidance behavior and negative affect. It may be that in the present model, the lateral portion of the OFC_med_lat layer exerting behavioral control over no-go responding may be partially serving the functions of the ACC. By this account, the lateral OFC would still maintain working memory information about recent negative reinforcement values (as OFC_ctxt does in our model), but this would then bias ACC activity when needed for behavioral control. This is supported by the existence of projections from the OFC to the ACC (Cavada et al., 2000), functional connectivity between these areas during risky decision making (M. X. Cohen, Heller, & Ranganath, 2005), and electrophysiological studies showing both lateral OFC and ACC signals during no-go trials in a go/no-go task (Bokura et al., 2001).

What determines DA firing? In the present work, we have also not addressed the important question of how reward and loss information is computed by systems upstream of midbrain DA neurons; instead, we simply assumed this function by externally increasing and decreasing simulated DA levels during positive reinforcement and negative reinforcement and then examined their effects on learning in the BG-OFC network. In parallel work, how interactions between distinct nuclei of the amygdala and ventral striatum can learn to associate stimuli with affective states and drive dopaminergic firing in the midbrain is being investigated (O'Reilly, Frank, Hazy, & Watz, 2005; see also Brown et al., 1999). This work provides a biologically explicit mechanism for the widely acknowledged relationship between the firing patterns of DA neurons during conditioning paradigms and those predicted by the abstract mathematical *temporal differences* reinforcement learning algorithm (Schultz et al., 1997; Sutton, 1988). This model produces DA firing patterns that are consistent with formal learning theory (e.g., in blocking paradigms; Kamin, 1968), as has been confirmed in DA recording experiments (Waelti, Dickinson, & Schultz, 2001). Future work will investigate the implications of this more ecologically valid computation of DA firing on decision-making processes within the BG and OFC.

More complex decision tasks: Dorsolateral PFC and BG circuits. We have also simplified the mechanisms by which information gets updated into working memory, in that all positive and negative experiences are updated and subsequently maintained. Researchers have argued that circuits linking the striatum with the (dorsolateral) PFC are involved in gating this updating function, allowing only task-relevant information to be maintained while distracting information is ignored (Frank et al., 2001; Hazy, Frank, & O'Reilly, in press; O'Reilly & Frank, 2006). This striatal gating function is further thought to be reinforced by dynamic changes in DA that support go and no-go BG firing to update (or not) prefrontal representations in an analogous fashion to the response selection circuits (Frank, 2005; O'Reilly & Frank, 2006). The current simulations did not include this striatal gating function of working memory updating because (a) the tasks used did not include distracting information, and (b) the encoding of reward magnitude information by the amygdala provides a similar, albeit more graded, gating function of OFC representations. Nevertheless, a BG gating function for selectively updating some dorsolateral PFC working memory representations, while continuing to robustly maintain others, may be critical for more cognitively demanding decision-making tasks involving planning and evaluation of if–then scenarios (Frank et al., 2001; O'Reilly & Frank, 2006).

Medial–lateral OFC versus right–left hemisphere. The medial and lateral OFC distinction of processing rewards and punishment that we chose to model remains controversial. Although several studies have provided evidence for this distinction, other theories have suggested a hemispheric specialization for processing positive and negative information. For example, using resting level electroencephalograph data (Davidson, 1992) and functional MRI data (e.g., Davidson & Irwin, 1999), Davidson and colleagues have provided a wealth of evidence suggesting that the left hemisphere is preferentially involved in emotions associated with approach, whereas the right hemisphere is involved in the expression of negative emotions. In addition, others have found increased depressive symptoms in stroke patients with damage to the left hemisphere (e.g., Robinson & Starkstein, 1989; Robinson & Szetela, 1981). Although these studies have not specifically investigated reward and punishment processing in the frontal cortex, they nonetheless are important to consider in theories of the neural mechanisms of decision making.

Further empirical testing of these competing processing distinctions is needed. Although we chose to use a medial–lateral distinction for reward–punishment outcome processing, the crucial principle implemented in the model is that reward and punishment magnitudes are represented in a distributed fashion in the OFC. Theoretically, a strict anatomical division for positive and negative values might not be necessary, as long as these can be differentiated within distributed neural populations. In this scenario, positive reward representations could come to be associated with striatal go signals by virtue of these having been paired together (i.e., via Hebbian learning), whereas negative representations would be consistently associated with BG no-go signals. Thus, disconfirming evidence for a medial–lateral distinction should not discredit the general mechanisms proposed in the model.

Serotonin. Our model is also incomplete at the level of neurotransmitters. In particular, some have suggested that whereas positive reinforcement learning depends on DA bursts, the magnitude of DA dips might be too small to functionally support negative learning (Bayer & Glimcher, 2005). Further, others have argued that serotonin (5-HT) is more likely to support negative learning, via opponent processes with DA (Daw, Kakade, & Dayan, 2002; Deakin, 1983). Our position on this issue is that (a) 5-HT may indeed be involved in negative reinforcement learning, but (b) this effect is likely to be mediated within the PFC (specifically, the OFC) rather than the BG, and (c) low levels of DA remain critical for BG-mediated negative reinforcement learning.

In support of a role for 5-HT in negative reinforcement no-go learning, elevated 5-HT levels are associated with enhanced error processing and harm avoidance behavior (Falgatter et al., 2004; Mazzanti et al., 1998). Acute depletion of 5-HT in healthy volunteers impairs reversal learning (Murphy, Smith, Cowen, Robbins, & Sahakian, 2002) and increases impulsive choice (Evers et al., 2005; Walderhaug et al., 2002), characteristic markers of impaired no-go learning and behavior. Reversal learning deficits have recently been observed in monkeys with 5-HT depletion selective to the PFC (Clarke et al., 2005). Others have shown that OFC lesions increase compulsive responding and that this is reversed by a selective serotonin reuptake inhibitor (Joel, Doljansky, Roz, & Rehavi, 2005). Taken together, these results suggest that 5-HT is involved in negative reinforcement learning and avoidance behavior but that this effect is probably mediated within the OFC-PFC rather than the BG. It could be that the DA dips that reinforce lateral OFC representations in the current model are additionally supported by increases in 5-HT, which promote active maintenance of recently negative reinforcement value.

We suggest that low DA levels remain critical for no-go learning that is mediated within the BG. First, as argued previously, although the magnitude of DA dips is smaller than that of bursts, this is likely compensated for by the far greater sensitivity of D_2 (compared with D_1) receptors to low DA levels (Creese, Sibley, Hamblin, & Leff, 1983; Goto & Grace, 2005). Thus, smaller DA dips would be sufficient to support no-go learning via D_2 receptors, whereas larger phasic bursts would be needed to support go learning via D_1 receptors. Second, there is a growing body of evidence that low DA levels are critical for no-go learning in humans. Researchers have found that PD patients are better at no-go learning when they are not taking DA medication (Frank et al., 2004). Although this learning could have been supported by the 5-HT system (Bayer & Glimcher, 2005), this possibility is unlikely given that 5-HT levels are actually reduced in PD along with DA (Di Cara, Samuel, Salin, Kerkerian-Le Goff, & Daszuta, 2003; Doder, Rabiner, Turjanski, Lees, & Brooks, 2003). Moreover, several studies report no-go learning deficits in humans taking DA medications, especially those acting on the D_2 receptor (Cools et al., 2001, 2003; Frank & O'Reilly, in press; Frank et al., 2004; Mehta, Swainson, Ogilvie, Sahakian, & Robbins, 2000); according to our model, these medications would block the effects of DA dips needed to learn no go (Frank, 2005). Notably, none of these effects can be explained via serotonergic mechanisms: Given the unidirectional inhibition of 5-HT onto DA (and not vice versa; Daw et al., 2002), acute DA manipulation is unlikely to have substantial effects on 5-HT. Indeed, whereas 5-HT inhibits DA release (Kapur & Remington, 1996; potentially via 5-HT receptors in DA cells, Nocjar, Roth, & Pehek, 2002), there is no reciprocal

relationship for DA onto 5-HT (Adell & Artigas, 1999). Finally, although DA dips have been repeatedly observed when outcomes are worse than expected (e.g., Bayer & Glimcher, 2005; Hollerman & Schultz, 1998; Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004; Satoh et al., 2003; Schultz, 1999; Ungless, Magill, & Bolam, 2004), the predicted phasic increase in 5-HT during these negative prediction errors has yet to be observed (see Daw et al., 2002), and if it is observed, it is unclear how it would preferentially bias no-go learning in the BG.

Other psychological phenomena: Heuristics and biases. Finally, in addition to the above neurobiological data, our model is not currently able to simulate all findings from the heuristics and biases literature. Because this is a model that must train and learn from feedback over the course of several trials, it is unlikely that we would be able to simulate findings in which the availability bias, the anchoring and adjustment bias, or the representative bias have been shown (e.g., Tversky & Kahneman, 1974). Also, the types of problems used to elicit these biases require abstract concepts of probability and verbal labels for positive and negative events, which are the result of extensive experience with the world. Prospect theory (Kahneman & Tversky, 1979) has been very successful at explaining the above results by including decision weights that contribute to expected value computations, but the mechanisms that lead to the computations are not specified. We suspect that with extensive training and expanded capacity, our model would be able to reproduce many of the findings showing the use of processing heuristics and would give a mechanistic account of these phenomenon.

Conclusion

We extended a previous model of the BG-DA system in reinforcement learning to explore how it interacts with simulated orbitofrontal and amygdalar contributions. We showed that the primitive BG model learns to slowly integrate reinforcement values over multiple trials of experience and is thus suitable for learning to make choices based on their frequencies of positive versus negative outcomes. Orbitofrontal contributions are necessary, however, when reinforcement information changes rapidly or when magnitude of gains and losses is more relevant than their frequency of occurrence. We outlined several novel behavioral and neural predictions of the augmented model and are hopeful that whether these are confirmed or falsified, they will lead to theoretical and practical advances.

References

Adell, A., & Artigas, F. (1999). Regulation of the release of 5-hydroxytryptamine in the median raphe nucleus of the rat by catecholaminergic afferents. *European Journal of Neuroscience, 11,* 2305–2311.

Alexander, G. E., & Crutcher, M. D. (1990a). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neuroscience, 13,* 266–271.

Alexander, G. E., & Crutcher, M. D. (1990b). Preparation for movement: Neural representations of intended direction in three motor areas of the monkey. *Journal of Neurophysiology, 64,* 133–150.

Alexander, G., Crutcher, M., & DeLong, M. (1990). Basal ganglia–thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. In H. Uylings, C. Van Eden, J. De Bruin, M. Corner, & M. Feenstra (Eds.), *The prefrontal cortex: Its structure, function, and pathology* (pp. 119–146). Amsterdam: Elsevier.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience, 6,* 196–202.

Annett, L. E., McGregor, A., & Robbins, T. W. (1989). The effects of ibotenic acid lesions of the nucleus accumbens on spatial learning and extinction in the rat. *Behavioral Brain Research, 31,* 231–242.

Arana, F. S., Parkinson, J. A., Hinton, E., Holland, A. J., Owen, A. M., & Roberts, A. C. (2003). Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *Journal of Neuroscience, 23,* 9632–9638.

Atallah, H. E., Frank, M. J., & O'Reilly, R. C. (2004). Hippocampus, cortex and basal ganglia: Insights from computational models of complementary learning systems. *Neurobiology of Learning and Memory, 82,* 253–267.

Aubert, I., Ghorayeb, I., Normand, E., & Bloch, B. (2000). Phenotypical characterization of the neurons expressing the D1 and D2 dopamine receptors in the monkey striatum. *Journal of Comparative Neurology, 418,* 22–32.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37,* 407–419.

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

Baxter, M. G., Parker, A., Lindner, C. C., Izquierdo, A. D., & Murray, E. A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *Journal of Neuroscience, 20,* 4311–4319.

Bayer, H. M. (2004). *A role for the substantia nigra in learning and motor control.* Unpublished doctoral thesis, New York University, New York.

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

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior, 52,* 336–372.

Bechara, A., Damasio, H., Damasio, A. R., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50,* 7–15.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience, 18,* 428–437.

Bechara, A., Dolan, S., & Hindes, A. (2002). Decision-making and addiction (Part II): Myopia for the future or hypersensitivity to reward? *Neuropsychologia, 40,* 1690–1705.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex, 6,* 215–225.

Beiser, D. G., & Houk, J. C. (1998). Model of cortical–basal ganglionic processing: Encoding the serial order of sensory events. *Journal of Neurophysiology, 79,* 3168–3188.

Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction, 96,* 73–86.

Bokura, H., Yamaguchi, S., & Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a go/no-go task. *Clinical Neurophysiology, 112,* 2224–2232.

Bolla, K. I., Eldreth, D. A., London, E. D., Kiehl, K. A., Mouratidis, M., Contoreggi, C., et al. (2003). Orbitofrontal cortex dysfunction in abstinent cocaine abusers performing a decision-making task. *NeuroImage, 19,* 1085–1094.

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. W., 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.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment, 14,* 253–262.

Calabresi, P., Saiardi, A., Pisani, A., Baik, J. H., Centonze, D., Mercuri, N. B., et al. (1997). Abnormal synaptic plasticity in the striatum of mice lacking dopamine D2 receptors. *Journal of Neuroscience, 17,* 4536–4544.

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology, 363,* 615–641.

Cavada, C., Company, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suarez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex: A review. *Cerebral Cortex, 10,* 220–242.

Centonze, D., Picconi, B., Gubellini, P., Bernardi, G., & Calabresi, P. (2001). Dopaminergic control of synaptic plasticity in the dorsal striatum. *European Journal of Neuroscience, 13,* 1071–1077.

Chudasama, Y., & Robbins, T. W. (2003). Dissociable contributions of the orbitofrontal and infralimbic cortex to Pavlovian autoshaping and discrimination reversal learning: Further evidence for the functional heterogeneity of the rodent frontal cortex. *Journal of Neuroscience, 23,* 8771–8780.

Clarke, H. F., Walker, S. C., Crofts, H. S., Dalley, J. W., Robbins, T. W., & Roberts, A. C. (2005). Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. *Journal of Neuroscience, 25,* 532–538.

Claus, E. D. (2005). *Rewards and punishments as determinants of behavior: Using personality to understand mechanisms of decision making.* Unpublished master's thesis, University of Colorado at Boulder.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychological Review, 97,* 332–361.

Cohen, M. X., Heller, A. S., & Ranganath, C. (2005). Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Brain Research, 23,* 61–70.

Cools, R. (2005). Dopaminergic modulation of cognitive function—Implications for l-DOPA treatment in Parkinson's disease. *Neuroscience and Biobehavioral Reviews, 30,* 1–23.

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral Cortex, 11,* 1136–1143.

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2003). L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinson's disease. *Neuropsychologia, 41,* 1431–1441.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22,* 4563–4567.

Creese, I., Sibley, D. R., Hamblin, M. W., & Leff, S. E. (1983). The classification of dopamine receptors: Relationship to radioligand binding. *Annual Review of Neuroscience, 6,* 43–71.

Crutcher, M. D., & Alexander, G. E. (1990). Movement-related neuronal activity selectively coding either direction or muscle pattern in three motor areas of the monkey. *Journal of Neurophysiology, 64,* 151–163.

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

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition, 20,* 125–151.

Davidson, R. J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences, 3,* 11–21.

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

Deakin, J. F. W. (1983). Roles of brain serotonergic neurons in escape,

avoidance and other behaviors. *Journal of Psychopharmacology, 43,* 563–577.

Deco, G., & Rolls, E. T. (2005). Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. *Cerebral Cortex, 15,* 15–30.

Delgado, M. R., Miller, M. M., Inati, S., & Phelps, E. A. (2005). An fMRI study of reward-related probability learning. *NeuroImage, 24,* 862–873.

Di Cara, B., Samuel, D., Salin, P., Kerkerian-Le Goff, L., & Daszuta, A. (2003). Serotonergic regulation of the GABAergic transmission in the rat basal ganglia. *Synapse, 50,* 144–150.

Doder, M., Rabiner, E. A., Turjanski, N., Lees, A. J., & Brooks, D. J. (2003). Tremor in Parkinson's disease and serotonergic dysfunction: An 11C-WAY 100635 PET study. *Neurology, 60,* 601–605.

Dodd, M.L., Klos, K.J., Bower, J.H., Geda, Y.E., Josephs, K.A. & Ahls-kog, J.E. (2005). Pathological gambling caused by drugs used to treat Parkinson disease. *Archives of Neurology, 62,* 1377—1381

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology, 83,* 1733–1750.

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, 10,* 308–317.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211.

Ernst, M., Bolla, K., Mouratidis, M., Contoreggi, C., Matochik, J. A., Kurian, V., et al. (2002). Decision-making in a risk-taking task: A PET study. *Neuropsychopharmacology, 26,* 682–691.

Estes, W. K. (1961). A descriptive approach to the dynamics of choice behavior. *Behavioral Science, 6,* 177–184.

Evers, E. A., Cools, R., Clark, L., van der Veen, F. M., Jolles, J., Sahakian, B. J., & Robbins, T. W. (2005). Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. *Neuropsychopharmacology, 30,* 1138–1147.

Falgatter, A. J., Herrmann, M. J., Roemmler, J., Ehlis, A. C., Wagener, A., Heidrich, A., et al. (2004). Allelic variation of serotonin transporter function modulates the brain electrical response for error processing. *Neuropsychopharmacology, 29,* 1506–1511.

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain, 126,* 1830–1837.

Finch, D. M. (1999). Plasticity of responses to synaptic inputs in rat ventral striatal neurons after repeated administration of the dopamine D2 antagonist raclopride. *Synapse, 31,* 297–301.

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. (in press). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks* (2006 special issue on the neurobiology of decision making).

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. (in press). A mechanistic account of striatal dopamine function in cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience.*

Frank, M. J., O'Reilly, R. C., & Curran, T. (in press). When memory fails, intuition reigns: Midazolam enhances implicit inference in humans. *Psychological Science.*

Frank, M. J., Rudy, J. W., & O'Reilly, R. C. (2003). Transitivity, flexibility, conjunctive representations and the hippocampus: II. A computational analysis. *Hippocampus, 13,* 341–354.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004, December 10). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science, 306,* 1940–1943.

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron, 47,* 495–501.

Fuster, J. M. (1989). *The prefrontal cortex: Anatomy, physiology and neuropsychology of the frontal lobe.* New York: Raven Press.

Fuster, J. M. (1997). *The prefrontal cortex: Anatomy, physiology and neuropsychology of the frontal lobe* (3rd ed.). New York: Lippincott-Raven.

Gallagher, M., McMahan, R. W., & Schoenbaum, G. (1999). Orbitofrontal cortex and representation of incentive value in associate learning. *Journal of Neuroscience, 19,* 6610–6614.

Gerfen, C. R. (1992). The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. *Annual Review of Neuroscience, 15,* 285–320.

Gerfen, C. R. (2000). Molecular effects of dopamine on striatal projection pathways. *Trends in Neurosciences, 23,* S64–S70.

Goldman-Rakic, P. S. (1995). Architecture of the prefrontal cortex and the central executive. *Annals of the New York Academy of Sciences, 769,* 71–83.

Gonzalez, C., Danda, J., Koshino, H., & Just, M. (2005). The framing effect and risky decisions: Examining cognitive functions with fMRI. *Journal of Economic Psychology, 26,* 1–20.

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

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22,* 10829–10837.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003, August 22). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science, 301,* 1104–1107.

Grant, S., Contoreggi, C., & London, E. D. (2000). Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia, 38,* 1180–1187.

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.

Haber, S., Kunishio, K., Mizobuchi, M., & Lynd-Balta, E. (1995). The orbital and medial prefrontal circuit through the primate basal ganglia. *Journal of Neuroscience, 15,* 4851–4867.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (in press). Banishing the homunculus: Making working memory work. *Neuroscience.*

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 Ca2+ conductance. *Journal of Neuroscience, 17,* 3334–3342.

Hernandez-Lopez, S., Tkatch, T., Perez-Garci, E., Galarraga, E., Bargas, J., Hamm, H., & Surmeier, D. J. (2000). D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca2+ currents and excitability via a novel PLCβ1-IP3-calcineurin-signaling cascade. *Journal of Neuroscience, 20,* 8987–8995.

Hikosaka, K., & Watanabe, M. (2000). Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cerebral Cortex, 10,* 263–271.

Hikosaka, K., & Watanabe, M. (2004). Long- and short-range reward expectancy in the primate orbitofrontal cortex. *European Journal of Neuroscience, 19,* 1046–1054.

Hikosaka, O. (1994). Role of basal ganglia in control of innate movements, learned behaviour and cognition. In G. Percheron, J. McKenzie, & J. Feger (Eds.), *The basal ganglia IV: New ideas and data on structure and function* (pp. 589–596). New York: Plenum Press.

Hikosaka, O. (1998). Neural systems for control of voluntary action–A hypothesis. *Advances in Biophysics, 35,* 81–102.

Hinson, J. M., & Staddon, J. E. (1983). Matching, maximizing and hill-

climbing. *Journal of the Experimental Analysis of Behavior, 40,* 321–331.

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

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience, 1,* 304–309.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.

Horn, N. R., Dolan, M., Elliott, R., Deakin, J. F., & Woodruff, P. W. (2003). Response inhibition and impulsivity: An fMRI study. *Neuropsychologia, 41,* 1959–1966.

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, 5,* 95–110.

Iversen, S., & Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Experimental Brain Research, 11,* 376–386.

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience, 24,* 7540–7548.

Joel, D., Doljansky, J., Roz, N., & Rehavi, M. (2005). Role of the orbital cortex and of the serotonergic system in a rat model of obsessive compulsive disorder. *Neuroscience, 130,* 25–36.

Joel, D., & Weiner, I. (1999). Striatal contention scheduling and the split circuit scheme of basal ganglia–thalamocortical circuitry: From anatomy to behaviour. In R. Miller & J. R. Wickens (Eds.), *Conceptual advances in brain research: Brain dynamics and the striatal complex* (pp. 209–236). Amsterdam: Harwood Academic.

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., & Graybiel, A. M. (1999, November 26). Building neural representations of habits. *Science, 286,* 1745–1749.

Johansen, J. P., & Fields, H. L. (2004). Glutamatergic activation of anterior cingulate cortex produces an aversive teaching signal. *Nature Neuroscience, 7,* 398–403.

Jones, B., & Mishkin, M. (1972). Limbic lesions and the problem of stimulus–reinforcement associations. *Experimental Neurology, 36,* 362–377.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior* (pp. 9–31). Miami, FL: University of Miami Press.

Kapur, S., & Remington, G. (1996). Serotonin–dopamine interaction and its relevance to schizophrenia. *American Journal of Psychiatry, 153,* 466–476.

Kemp, J. M., & Powell, T. P. (1970). The cortico-striate projections in the monkey. *Brain, 93,* 525–546.

Kerr, J. N., & Wickens, J. R. (2001). Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. *Journal of Neurophysiology, 85,* 117–124.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996, September 6). A neostriatal habit learning system in humans. *Science, 273,* 1399–1402.

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: Characterization with rapid event-related fMRI. *NeuroImage, 18,* 263–272.

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Progress in Neurobiology, 72,* 341–372.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23,* 155–184.

Lindvall, O., Bjorklund, A., & Divac, I. (1978). Organization of the catecholamine neurons projecting to the frontal cortex in the rat. *Brain Research, 142,* 1–24.

Major, G., & Tank, D. (2004). Persistent neural activity: Prevalence and mechanisms. *Current Opinion in Neurobiology, 14,* 675–684.

Markowitsch, H. J., Vandekerckhovel, M. M., Lanfermann, H., & Russ, M. O. (2003). Engagement of lateral and medial prefrontal areas in the ecphory of sad and happy autobiographical memories. *Cortex, 39,* 643–665.

Mazzanti, C. M., Lappalainen, J., Long, J. C., Bengel, D., Naukkarinen, H., Eggert, M., et al. (1998). Role of the serotonin transporter promoter polymorphism in anxiety-related traits. *Archives of General Psychiatry, 55,* 936–940.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004, October 15). Separate neural systems value immediate and delayed rewards. *Science, 306,* 503–507.

McCoy, A. N., & Platt, M. L. (2005). Expectations and outcomes: Decision making in the primate brain. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology,* 191, 201–211.

Mehta, M. A., Swainson, R., Ogilvie, A. D., Sahakian, B. J., & Robbins, T. W. (2000). Improved short-term spatial memory but impaired reversal learning following the dopamine D2 agonist bromocriptine in human volunteers. *Psychopharmacology, 159,* 10–20.

Mesulam, M.-M. (2000). Paralimbic (mesocortical) areas. In M.-M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (pp. 49–54). New York: Oxford University Press.

Milham, M. P., Crowley, T. J., Thompson, L. L., Raymond, K. M., Claus, E. D., & Banich, M. T. (2006). *Reduction in orbitofrontal brain volume in adolescents with severe substance and conduct problems.* Manuscript submitted for publication.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16,* 5154–5167.

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

Mobini, S., Body, S., Ho, M. Y., Bradshaw, C. M., Szabadi, E., Deakin, J. F., & Anderson, I. M. (2002). Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology, 160,* 290–298.

Morris, G., Arkadir, D., Nevet, A., Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron, 43,* 133–143.

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.

Muller, U., Wachter, T., Barthel, H., Reuter, M., & von Cramon, D. Y. (2000). Striatal [123I]beta-CIT SPECT and prefrontal cognitive functions in Parkinson's disease. *Journal of Neural Transmission, 107,* 303–319.

Murphy, F. C., Smith, K. A., Cowen, P. J., Robbins, T. W., & Sahakian, B. J. (2002). The effects of tryptophan depletion on cognitive and affective processing in healthy volunteers. *Psychopharmacology, 163,* 42–53.

Nishi, A., Snyder, G. L., & Greengard, P. (1997). Bidirectional regulation of DARPP-32 phosphorylation by dopamine. *Journal of Neuroscience, 17,* 8147–8155.

Nocjar, C., Roth, B. L., & Pehek, E. A. (2002). Localization of 5-HT(2A) receptors on dopamine cells in subnuclei of the midbrain A10 cell group. *Neuroscience, 111,* 163–176.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum Press.

O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003).

Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *Journal of Neuroscience, 23,* 7931–7939.

O'Doherty, J., Kringelback, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment in the human orbitofrontal cortex. *Nature Neuroscience, 4,* 95–102.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology, 15,* 267–273.

Onn, S. P., West, A. R., & Grace, A. A. (2000). Dopamine-mediated regulation of striatal neuronal and network interactions. *Trends in Neuroscience, 23*(Suppl.1), S48–S56.

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

O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences, 2,* 455–462.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the frontal cortex and basal ganglia. *Neural Computation, 18,* 283–328.

O'Reilly, R. C., Frank, M. J., Hazy, T. E., & Watz, B. (2005). *Rewards are timeless: The primary value and learned value (PVLV) Pavlovian learning algorithm.* Manuscript submitted for publication.

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

O'Reilly, R. C., Noelle, D., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex, 12,* 246–257.

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

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,* 6235–6242.

Parkinson, J. A., Dalley, J. W., Cardinal, R. N., Bamford, A., Fehnert, B., Lachenal, G., et al. (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: Implications for mesoaccumbens dopamine function. *Behavioral Brain Research, 137,* 149–163.

Pasupathy, A., & Miller, E. K. (2005, February 24). Different time courses for learning-related activity in the prefrontal cortex and striatum. *Nature, 433,* 873–876.

Paulus, M. P., Rogalsky, C., Simmons, A., Feinstein, J. S., & Stein, M. B. (2003). Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. *NeuroImage, 19,* 1439–1448.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Peters, E., & Slovic, P. (2000). The springs of action: Affective and analytical information processing in choice. *Personality and Social Psychology Bulletin, 26,* 1465–1475.

Pickens, C. L., Saddoris, M. P., Gallagher, M., & Holland, P. C. (2005). Orbitofrontal lesions impair use of cue–outcome associations in a devaluation task. *Behavioral Neuroscience, 119,* 317–322.

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., & Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *Journal of Neuroscience, 23,* 11078–11084.

Platt, M. L. (2002). Neural correlates of decisions. *Current Opinion in Neurobiology, 12,* 141–148.

Plous, S. (1993). *The psychology of judgment and decision making.* New York: McGraw-Hill.

Remijnse, P. L., Nielen, M. M. A., Uylings, H. B. M., & Veltman, D. J. (2005). Neural correlates of a reversal learning task with an affectively

neutral baseline: An event-related fMRI study. *NeuroImage, 26,* 609–618.

Remy, P., Jackson, P. L., & Ribeiro, M. J. (2000). Relationships between cognitive deficits and dopaminergic function in the striatum of Parkinson's disease patients: A PET study. *Neurology, 54* (Suppl. 3), A372.

Robertson, G. S., Vincent, S. R., & Fibiger, H. C. (1992). D1 and D2 dopamine receptors differentially regulate c-fos expression in striatonigral and striatopallidal neurons. *Neuroscience, 49,* 285–296.

Robinson, R. G., & Starkstein, S. E. (1989). Mood disorders following stroke: New findings and future directions. *Journal of Geriatric Psychiatry, 22,* 1–15.

Robinson, R. G., & Szetela, B. (1981). Mood change following left hemispheric brain injury. *Annals of Neurology, 9,* 447–453.

Roesch, M. R., & Olson, C. R. (2004, April 9). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science, 304,* 307–310.

Rogers, R. D., Owen, A. M., Middleton, H. C., Williams, E. J., Pickard, J. D., Sahakian, B. J., & Robbins, T. W. (1999). Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex. *Journal of Neuroscience, 19,* 9029–9038.

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

Rolls, E. T. (1999). *The brain and emotion.* Oxford University Press.

Rolls, E. T. (2004). Convergence of sensory systems in the orbitofrontal cortex in primates and brain design for emotion. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology, 281A,* 1212–1225.

Rolls, E. T., & Kringelbach, M. L. (2003). Different representations of pleasant and unpleasant odours in the human brain. *European Journal of Neuroscience, 18,* 695–703.

Rosenkilde, C. E., Bauer, R. H., & Fuster, J. M. (1981). Single cell activity in ventral prefrontal cortex of behaving monkeys. *Brain Research, 209,* 375–394.

Rubchinsky, L. L., Kopell, N., & Sigvardt, K. A. (2003). Modeling facilitation and inhibition of competing motor programs in basal ganglia subthalamic nucleus–pallidal circuits. *Proceedings of the National Academy of Sciences, USA, 100,* 14427–14432.

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.

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., Chiba, A. A., & Gallagher, M. (2000). Changes in functional connectivity in orbitofrontal cortex and basolateral amygdala during learning and reversal training. *Journal of Neuroscience, 20,* 5179–5189.

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

Schoenbaum, G., & Setlow, B. (2001). Integrating orbitofrontal cortex into prefrontal theory: Common processing themes across species and subdivisions. *Learning and Memory, 8,* 134–147.

Schoenbaum, G., & Setlow, B. (2003). Lesions of nucleus accumbens disrupt learning about aversive outcomes. *Journal of Neuroscience, 23,* 9833–9841.

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

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,* 1–27.

Schultz, W. (1999). The reward signal of midbrain dopamine neurons. *News in Physiological Sciences, 14,* 249–255.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron, 36,* 241–263.

Schultz, W., Dayan, P., & Montague, P. R. (1997, March 14). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Seamans, J. K., Durstewitz, D., Christie, B. R., Stevens, C. F., & Sejnowski, T. J. (2001). Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. *Proceedings of the National Academy of Sciences, USA, 98,* 301–306.

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.

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

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making, 15,* 233–250.

Shima, K., & Tanji, J. (1998, November 13). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science, 282,* 1335–1338.

Shiv, B., Loewenstein, G., & Bechara, A. (2005). The dark side of emotion in decision-making: When individuals with decreased emotional reactions make more advantageous decisions. *Cognitive Brain Research, 23,* 85–92.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain, 127,* 851–859.

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C., & Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: From pleasure to aversion. *Brain, 124*(Pt. 9), 1720–1733.

Smith-Roe, S. L., & Kelley, A. E. (2000). Coincident activation of NMDA and dopamine D1 receptors within the nucleus accumbens core is required for appetitive instrumental learning. *Journal of Neuroscience, 22,* 7737–7742.

Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin & Review, 11,* 742–747.

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

Swainson, R., Rogers, R. D., Sahakian, B. J., Summers, B. A., Polkey, C. E., & Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia, 38,* 596–612.

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005, March 11). Adaptive coding of reward value by dopamine neurons. *Science, 307,* 1642–1645.

Tremblay, L., & Schultz, W. (1999, April 22). Relative reward preference in primate orbitofrontal cortex. *Nature, 398,* 704–708.

Tremblay, L., & Schultz, W. (2000). Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *Journal of Neurophysiology, 83,* 1877–1885.

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Tversky, A., & Kahneman, D. (1981, January 30). The framing of decisions and the psychology of choice. *Science, 211,* 453–458.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* S251–S278.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference dependent model. *Quarterly Journal of Economics, 107,* 1039–1061.

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004, March 26). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science, 303,* 2040–2042.

Ursu, S., & Carter, C. S. (2005). Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: Implications for neuroimaging studies of decision-making. *Brain Research, 23,* 51–60.

Volkow, N. D., Chang, L., Wang, G. J., Fowler, J. S., Ding, Y. S., Sedler, M., et al. (2001). Low level of brain dopamine D2 receptors in methamphetamine abusers: Association with metabolism in the orbitofrontal cortex. *American Journal of Psychiatry, 158,* 2015–2021.

Volkow, N. D., Fowler, J. S., & Wang, G. J. (2003). The addicted human brain: Insights from imaging studies. *Journal of Clinical Investigation, 111,* 1444–1451.

Waelti, P., Dickinson, A., & Schultz, W. (2001, July 5). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412,* 43–48.

Wagar, B. M., & Thagard, P. (2004). Spiking Phineas Gage: A neurocomputational theory of cognitive–affective integration in decision making. *Psychological Review, 111,* 67–79.

Walderhaug, E., Lunde, H., Nordvik, J. E., Landro, N. I. Refsum, H., & Magnusson, A. (2002). Lowering of serotonin by rapid tryptophan depletion increases impulsiveness in normal individuals. *Psychopharmacology, 164,* 385–391.

Wallis, J. D., & Miller, E. K. (2003). From rule to response: Neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology, 90,* 1790–1806.

Wang, J., & O'Donnell, P. (2001). D1 dopamine receptors potentiate NMDA-mediated excitability increase in layer V prefrontal cortical pyramidal neurons. *Cerebral Cortex, 11,* 452–462.

Watanabe, K., & Hikosaka, O. (2005). Immediate changes in anticipatory activity of caudate neurons associated with reversal of position–reward contingency. *Journal of Neurophysiology, 94,* 1879–1887.

Wickens, J. (1997). Basal ganglia: Structure and computations. *Network: Computation in Neural Systems, 8,* R77–R109.

Wise, S. P., Murray, E. A., & Gerfen, C. R. (1996). The frontal cortex–basal ganglia system in primates. *Critical Reviews in Neurobiology, 10,* 317–356.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111,* 931–959.

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

Appendix

Implementational Details

The model is implemented through the use of a subset of the Leabra framework (O'Reilly, 1998; O'Reilly & Munakata, 2000). Leabra uses point neurons with excitatory, inhibitory, and leak conductances contributing to an integrated membrane potential, which is then thresholded and transformed via an $x/(x + 1)$ sigmoidal function to produce a rate code output communicated to other units. (Discrete spiking can also be used, but it produces noisier results.) Each layer uses a k-winners-take-all (kWTA) function that computes an inhibitory conductance that keeps roughly the k most active units above firing threshold and keeps the rest below threshold. Units learn according to a combination of Hebbian and error-driven learning, with the latter computed with the generalized recirculation algorithm (O'Reilly, 1996).

The membrane potential V_m is updated as a function of ionic conductances g with reversal (driving) potentials E as follows:

$$\frac{dV_m(t)}{dt} = \tau \Sigma_c g_c(t) \bar{g}_c [E_c - V_m(t)], \tag{1}$$

with three channels (c) corresponding to excitatory input e, leak current l, and inhibitory input i. 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 $\bar{g}_c$ that controls the relative influence of the different conductances. The equilibrium potential can be written in a simplified form by setting the excitatory driving potential (E_e) to 1 and the leak and inhibitory driving potentials (E_l and E_i) to 0:

$$V_m^{\infty} = \frac{g_e \bar{g}_e}{g_e \bar{g}_e + g_l \bar{g}_l + g_i \bar{g}_i}, \tag{2}$$

which shows that the neuron is computing a balance between excitation and the opposing forces of leak and inhibition. This equilibrium form of the equation can be understood in terms of a Bayesian decision-making framework (O'Reilly & Munakata, 2000).

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

$$\eta_j = g_e(t) = \ <x_i w_{ij}> \ = \frac{1}{n} \Sigma_i x_i w_{ij}. \tag{3}$$

The inhibitory conductance is computed via the kWTA function described in the next section, and leak is a constant. In the equation above, x_i is a sending activation value; w_{ij} is a synaptic weight from unit i to unit j, and angle brackets indicate normalized average.

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

$$y_j(t) = \frac{1}{\left(1 + \frac{1}{\gamma [V_m(t) - \Theta]_+}\right)}, \tag{4}$$

where γ is a gain parameter and $[x]_+$ is a threshold function that returns 0 if $x < 0$ and returns to x if $x > 0$. Note that if it returns 0, we assume $y_j(t) = 0$, to avoid dividing by 0. As it is, this function has a very sharp threshold, which interferes with graded learning mechanisms (e.g., gradient descent). To produce a less discontinuous deterministic function with a softer threshold, we convolved the function with a Gaussian noise kernel ($\mu = 0$, $\sigma = .005$), which reflects the intrinsic processing noise of biological neurons:

$$y_j(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-z^2/(2\sigma^2)} y_j(z - x) dz, \tag{5}$$

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 through the use of a numerical lookup table.

kWTA Inhibition Within Layers

Inhibition between layers (i.e., for GABAergic projections between BG layers) is achieved via simple unit inhibition, where the inhibitory current g_i for the unit is determined from the net input of the sending unit.

For within-layer inhibition, 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 for all units in the layer, such that the $k + 1$th most-excited unit within a layer is below its firing threshold, whereas the kth is above threshold. Activation dynamics similar to those produced by the kWTA function have been shown to result from simulated inhibitory interneurons that project both feedforward and feedback inhibition (O'Reilly & Munakata, 2000). Thus, although the kWTA function is somewhat biologically implausible in its implementation (e.g., requiring global information about activation states and using sorting mechanisms), it provides a computationally effective approximation to biologically plausible inhibitory dynamics.

The kWTA function is computed via a uniform level of inhibitory current for all units in the layer as follows:

$$g_i = g_{k+1}^{\Theta} + q(g_k^{\Theta} - g_{k+1}^{\Theta}), \tag{6}$$

where $0 < q < 1$ 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:

$$g_i^{\Theta} = \frac{g_e^* \bar{g}_e (E_e - \Theta) + g_l \bar{g}_l (E_l - \Theta)}{\Theta - E_i}, \tag{7}$$

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, 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. Thus, the inhibition is placed exactly to allow k units to be above threshold and the remainder below threshold. For this version, the q parameter is almost always .25, allowing the kth unit to be sufficiently above the inhibitory threshold.

Because the OFC layers have to represent differing magnitudes depending on the value of a particular outcome, they cannot use this standard kWTA inhibition. Instead, they require a more flexible function that allows the layer to sometimes have greater overall activity levels, depending on the excitatory input. The OFC layers therefore use the "AVG MAX PT" inhibition function, which simply puts the inhibition value between the average and maximum values for the layer. Specifically, g_k^{Θ} is the average g_i^{Θ} value for the layer, and g_{k+1}^{Θ} is the maximum g_i^{Θ} for the layer. Thus, greater excitatory input from the ABL results in greater overall OFC activity levels.

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 generalized recirculation algorithm (O'Reilly, 1996), which is functionally equivalent to the deterministic Boltzmann machine and contrastive Hebbian learning. The network settles in two phases, an expectation (minus) phase in which the network's actual output is produced and an outcome (plus) phase in which 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 (1982) normalization. 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

$$\Delta_{\text{hebb}} W_{ij} = x_i^+ y_j^+ - y_j^+ W_{ij} = y_j^+ (x_i^+ - w_{ij}), \tag{8}$$

and for error-driven learning using contrastive Hebbian learning is

$$\Delta_{\text{err}} w_{ij} = x_i^+ y_j^+ - x_i^- y_j^-, \tag{9}$$

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

$$\Delta_{\text{sberr}} w_{ij} = [\Delta_{\text{err}}]_+ (1 - w_{ij}) + [\Delta_{\text{err}}]_- w_{ij}. \tag{10}$$

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

$$\Delta w_{ij} = \epsilon [k_{\text{hebb}}(\Delta_{\text{hebb}}) + (1 - k_{\text{hebb}})(\Delta_{\text{sberr}})]. \tag{11}$$

OFC_ctxt Layer

The OFC_ctxt layer is a simple working memory context layer that reflects a combination of its own activity in the previous trial together with the activity of that in OFC_med_lat, as implemented in simple recurrent networks (Elman, 1990). Its units receive external input as a function of their previous activity and the activity of their sending units. Specifically, the external input u_ext equals

$$\text{updt.fm\_prv*u\_act\_p} + \text{updt.fm\_hid*su\_act\_p}, \tag{12}$$

where updt.fm_prv scales the previous plus-phase activity of that unit, u_act_p, and is set to .85, and updt.fm_hid scales the previous plus-phase activity of its corresponding sending unit, su_act_p, from OFC_med_lat by .60. This balance of parameters for maintaining previous states and updating new states allows the context layer to integrate reinforcement values over time, with larger values maintained more persistently than smaller values.

Simulated Lesions

In all cases of lesions in the model, we removed all of the units of the layer from processing. Thus, an OFC lesion corresponded to a model with no OFC effect on processing at all.