Annual Review of Neuroscience

Adaptive Cost-Benefit Control Fueled by Striatal Dopamine

Michael J. Frank

Department of Cognitive and Psychological Sciences and Carney Institute for Brain Science, Brown University, Providence, Rhode Island, USA; email: michael_frank@brown.edu

Annu. Rev. Neurosci. 2025. 48:1–22

The *Annual Review of Neuroscience* is online at neuro.annualreviews.org

https://doi.org/10.1146/annurev-neuro-112723-025228

Keywords

basal ganglia, striatum, dopamine, reinforcement learning, decision making

Abstract

The twenty-first century has brought forth a deluge of theories and data shedding light on the neural mechanisms of motivated behavior. Much of this progress has focused on dopaminergic dynamics, including their signaling properties (how do they vary with expectations and outcomes?) and their downstream impacts in target regions (how do they affect learning and behavior?). In parallel, the basal ganglia have been elevated from their original implication in motoric function to a canonical circuit facilitating the initiation, invigoration, and selection of actions across levels of abstraction, from motor to cognitive operations. This review considers how striatal D1 and D2 opponency allows animals to perform cost-benefit calculations across multiple scales: locally, whether to select a given action, and globally, whether to engage a particular corticostriatal circuit for guiding behavior. An emerging understanding of such functions reconciles seemingly conflicting data and has implications for neuroscience, psychology, behavioral economics, and artificial intelligence.

Contents

INTRODUCTION

Mammalian behaviors are incredibly adaptive to changes in the environment and fluctuating motivational drives. When fending for survival, the costs and benefits of alternative actions are quite different from those in a comfortable environment. Risky decisions can be adaptive if their potential benefits (relative to those of safer choices) outweigh the costs. Similarly, when opportunities for advancement present themselves, it may sometimes be worth the effort of working exceptionally hard for those gains, but sometimes it may be more productive to minimize (cognitive or physical) work and focus on leisure (Kool & Botvinick 2014). In this review I highlight decades of progress in theory and experimental data suggesting that the frontal cortex and basal ganglia (BG) interact to consider plausible actions for any given scenario, and to facilitate behaviors that differentially maximize benefits or minimize costs depending on the agent's motivational state and environmental context. This computation is fueled by heterogeneous dopamine (DA) dynamics across time scales and across corticostriatal circuits at multiple levels of abstraction, allowing animals to adaptively learn which circuits to engage and which actions to execute. Indeed, a long history of research implicates DA in the optimization of motivated behavior. While recent studies suggest heterogeneity in DA signals, I suggest that diverse DA dynamics embedded within striatal circuitry act in concert, supporting adaptive flexibility and motivated behavior tuned to task demands (Hamid et al. 2021, Jaskir & Frank 2023). This framework accords with a wide range of data that may have otherwise presented challenges for earlier ideas. I also review various open questions.

PHASIC DOPAMINE: REWARD PREDICTION ERRORS AND BEYOND

Arguably one of the greatest contributions of computational neuroscience was elucidating the ways in which dopaminergic signals correspond to reward prediction errors (RPEs), which are used in reinforcement learning (RL) (Montague et al. 1996, Schultz et al. 1997). Many experiments have supported the key predictions of these models, reviewed elsewhere (Watabe-Uchida et al. 2017). In brief, DA neurons respond in bursts to unpredicted rewards, and the magnitude of DA bursts is correlated with the magnitude of positive RPEs. When rewards are expected but not received, DA neurons exhibit transient dips in their activity, and the duration of the pause correlates with the magnitude of the negative prediction error (Bayer et al. 2007). Several causal manipulations across species have shown that these DA signals can induce RL in accordance with formal learning theory (Hamid et al. 2015, Stauffer et al. 2016, Steinberg et al. 2013), mediated by changes at the synaptic and intracellular levels in striatum (Lee et al. 2020, Scott & Frank 2023, Yagishita et al. 2014).

More recently, various findings have suggested heterogeneity in DA neuronal signaling, but where the combined population output still reflects RPE (Lee et al. 2024) or a distribution over RPEs (Dabney et al. 2020). Some studies show that DA signals surprising events even in the absence of any explicit reward value, allowing animals to learn arbitrary sensory-sensory associations (Sharpe et al. 2017). Yet, there remains an enormous amount of evidence that DA signals are valenced (i.e., rising when events are better than expected but decreasing when they are worse than expected). These opponent and symmetrical deflections are particularly prevalent in the striatum (Hart et al. 2014), albeit with some heterogeneity across subregions: As reviewed in the final section, valenced RPE signals are commonly observed across all major striatal subregions but are tailored to the computational functions of that region (Hamid et al. 2021, Mohebi et al. 2024, Parker et al. 2016, Tsutsui-Kimura et al. 2020). Moreover, as reviewed below, causal manipulations of DA and its downstream targets across striatum consistently implicate opponent effects on motivated behavior, with few exceptions. In contrast, DA signals in other target regions [e.g., amygdala (Lutas et al. 2019), medial shell of accumbens (de Jong et al. 2019), and prefrontal cortex (Abe et al. 2024)] show unvalenced signals of surprise, which may enhance associative learning about neutral and emotional memories.

BEYOND PHASIC TRANSIENTS: DOPAMINE RAMPS

DA concentrations also exhibit slower dynamics distinct from transient events. Whereas transient RPEs guide RL, I argue below that these slower signals relate to immediate motivational impacts of DA on behavior. Various studies have reported that DA slowly ramps over time as an animal makes progress toward a goal (Collins et al. 2016; Farrell et al. 2022; Goedhoop et al. 2023; Hamid et al. 2015, 2021; Howe et al. 2013; Syed et al. 2015). Such DA ramps are modulated by recent reward history (Hamid et al. 2015, Mohebi et al. 2019) and/or uncertainty (Mikhael et al. 2022) and are particularly apparent when the animal has agency to affect its outcomes (Collins et al. 2016, Goedhoop et al. 2023, Hamid et al. 2021, Syed et al. 2015). Many of these studies included direct measurements of striatal DA release, which can be locally controlled and can be dissociated from activity at midbrain cell bodies (Berke 2018, Mohebi et al. 2019).

Indeed, DA has long been implicated in motivated, reward-oriented behavior that is distinct and separable from learning (Berridge 2006, Salamone & Correa 2024). When striatal DA levels are elevated, humans and animals are increasingly likely to make choices that have higher risk, or require higher amounts of effort, if the potential payout is sufficiently high (Pagnier et al. 2024, Rutledge et al. 2015, Salamone & Correa 2024, St. Onge & Floresco 2009, Stopper et al. 2014, Zalocusky et al. 2016). Thus, DA ramps may provide local modulation of cost-benefit decision strategies as an animal makes progress toward a goal (just as one may engage in more effort toward the end of a race). Below, I argue for an opponent circuit mechanism by which striatal DA is sensitive to cost-benefit calculations that affect both learning and motivated decision making, and that these functions are intertwined.

BASAL GANGLIA CONTRIBUTIONS TO DECISION MAKING

While some studies have suggested that the BG primarily serve to invigorate ongoing motor output (Turner & Desmurget 2010), it is becoming increasingly clear that they can also causally guide which action is executed. The discussion in this section applies primarily to dorsal striatum;[1]

[1]Indeed, ventral striatum is particularly important for phenomena such as Pavlovian-to-instrumental transfer, whereby motor output is invigorated by a reward-predictive cue (Corbit & Balleine 2011).

for application of this framework to various striatal subregions, see the section titled A Mosaic of Cost-Benefit Control Across Corticostriatal Circuits.

Computational models of dynamic interactions can reconcile conflicting data in this regard. In these models, the premotor cortex proposes a small number of options that are relevant in the current sensory context [e.g., in proportion to their prior probabilities as seen empirically (Cisek & Kalaska 2005, Glaser et al. 2018)], and the striatum then accumulates the evidence for and against each one, ultimately disinhibiting thalamocortical activity associated with the best action (Bogacz & Gurney 2007; Calderon et al. 2022; Dunovan et al. 2019; Frank 2005, 2006; Ratcliff & Frank 2012). However, such disinhibition is only necessary for affecting choice when there is not already a dominant winning action in cortex. If cortical action proposals are sufficiently strong, they can reach threshold for execution without additional modulation by the BG and thalamus (Ashby et al. 2007, Frank & Claus 2006, Ratcliff & Frank 2012). Thus, the BG are only needed when there are conflicting sources of evidence for which action is best, requiring volitional drive (see below for more detail). Accordingly, Parkinson's patients show spared movements when they are compelled by sensory input but not when they require volitional initiation (Brown & Marsden 1988). Various other data implicate BG/DA specifically in motivated agency (Cockburn et al. 2014, Leotti & Delgado 2011, Stolz et al. 2020).

In perceptual decision making, a common paradigm requires animals to decide whether a noisy array of dots are moving primarily to the left or right (Roitman & Shadlen 2002). The amount of evidence is manipulated by varying the coherence of the dot motion. A rich literature has studied the neural basis for evidence accumulation in such paradigms, with vigorous debates about the causal contributions of any given brain region or neural population to choice. Indeed, Yartsev et al. (2018) articulated three characteristics necessary for a brain region to be causally involved in decision making and found that, to date, the dorsal striatum was the only region to satisfy all three criteria (striatal neurons reflected graded evidence accumulation, inactivating them impaired choice, and perturbing their activity had causal effects on choice throughout the accumulation process).

Critically, the striatal framework implies not only that decisions should reflect the evidence (coherence) but that this evidence should be reweighted according to the benefits (potential reward values) of each option. Indeed, when reward outcomes were manipulated in a random dot study, striatal neurons showed properties of evidence accumulation that combined the raw coherence with the reward values, ramping more steeply for those choices that have larger benefits (Doi et al. 2020). This study also confirmed in monkeys that perturbing striatal neurons causally altered decision making.

The above discussion suggests the BG are only needed for choice when there are conflicting sources of evidence on which action is best, and especially when this evidence is noisy and needs to be accumulated. But why should the BG then be involved in value-based choice without overt noise? Even here, regardless of whether values are learned within the task or based on, for example, food preferences, decision dynamics still conform to predictions from evidence accumulation models (Gluth et al. 2018, Ratcliff & Frank 2012, Thomas et al. 2021), and are modulated by BG activity (Frank et al. 2015, Pedersen & Frank 2020). According to the neural network models, action proposals are noisy (facilitating exploration), wavering sequentially over time until one of them is gated by BG-thalamocortical activity (Frank 2006, Ratcliff & Frank 2012; for video animations, see also **http://ski.clps.brown.edu/BGmodel_movies.html**). Indeed, during decision making, animals vacillate between covertly attending to the value of alternative choice options, with orbitofrontal spike rates reflecting the momentary value of covertly attended options (Ferro et al. 2024, Rich & Wallis 2016); decision vacillation is also observed in striatum (DePasquale et al. 2024). In humans, choice values are often not static and require accumulation of momentary

preferences—which can themselves vary with attention to different attributes (Busemeyer et al. 2019, Gluth et al. 2018, Thomas et al. 2021)—to make self-consistent choices. Notably, when benefits and costs of alternative choices were spatially presented to humans, participants sequentially attended to these attributes, leading to time-varying impacts of benefits and costs on evidence accumulation (Westbrook et al. 2020). Moreover, individual differences in, and manipulations of, striatal DA shifted the impact of attention to benefits and costs on evidence accumulation (Pagnier et al. 2024, Westbrook et al. 2020).

Striatal Opponency: From Go Versus No Go to Benefits Versus Costs

How do the BG differentially weight the benefits and costs of alternative actions? The notoriously intricate architecture of cortical-BG-thalamic circuits are most commonly described in terms of the direct and indirect pathways (Albin et al. 1989). According to the classical model, striatal medium spiny neurons (MSNs) in the direct pathway project to BG output and disinhibit thalamocortical populations related to the corresponding action. These neurons express D1 receptors and are excited by DA. Striatal MSNs in the indirect pathway project to the globus pallidus, which prevents action. These neurons express D2 receptors and are inhibited by DA (Gerfen & Surmeier 2011). Classical models thus suggest that DA promotes movement by activating the direct pathway and suppressing the indirect pathway (Albin et al. 1989). However, the simplest version of such a model cannot explain findings such as those showing that both striatal populations are coactivated during movement (Cui et al. 2013).

Dynamic neural network models of BG circuitry can resolve such discrepancies without eschewing opponency (Beeler et al. 2012, Calderon et al. 2022, Dunovan et al. 2019, Frank 2005, Frank & Claus 2006, Franklin & Frank 2015, Wiecki & Frank 2013). In these models, cortical action proposals are evaluated simultaneously by separate D1 and D2 populations coding for each action. While D1 neurons accumulate evidence for a given action, D2 neurons accumulate evidence against a particular action in part by suppressing its corresponding D1 population [via inhibitory collaterals from D2 to D1 neurons in the models and data (Beeler et al. 2012, Burke et al. 2017, Taverna et al. 2008)]. As such, activation of D2 neurons not only may suppress unwanted actions but also can promote selection of alternative actions by effectively disinhibiting D1 neurons associated with better actions. Such models thus show simultaneous activation of D1 and D2 neurons during action selection (Dunovan et al. 2019, Wiecki & Frank 2013) and where the balance is shifted to the D1 population coding for the to-be-selected action just prior to execution, as recently described empirically (Tang et al. 2024).

Moreover, in these models, DA activity dynamically modulates the relative engagement of each population, affecting both incremental RL and immediate motivated behavior. During outcomes, transient DA deflections (e.g., RPEs) induce activity-dependent synaptic plasticity, with DA bursts potentiating D1 neurons and DA dips potentiating D2 neurons (Beeler et al. 2012, Collins & Frank 2014, Dunovan et al. 2019, Frank 2005). Over the course of learning, D1 and D2 neurons come to represent the benefits and costs of alternative actions. During action selection, DA can also directly impact motivated choice behavior by modulating D1 versus D2 neuron activity levels and thus the relative impact of benefits versus costs.

A large body of evidence provides support for these opponent effects on motivated choice and learning across species. Regarding choice, we have already discussed the causal contribution of striatum to evidence accumulation (Doi et al. 2020, Yartsev et al. 2018). More recent studies have shown that D1 and D2 neurons contribute in opponent fashions: Silencing D1 cells impedes the ability to accumulate evidence for the contralateral action, whereas silencing D2 cells impedes the ability to accumulate evidence against that action (Bolkan et al. 2022). Critically, these striatal

manipulations had no influence on decision making when choices could be made based on raw immediate sensory information, or when animals were in a perseverative state insensitive to the current-trial evidence. These findings again suggest that the BG do not contribute when choices can be easily made based on immediate sensory input or when they are dictated by automated behavior based on strong corticocortical associations. Yet the BG are increasingly involved in modulating decision outputs during initial learning, when arbitrating between conflicting sources of evidence, and when decisions depend on the animal's internal (and potentially changing) motivational drives. Supporting the latter point, when deciding whether to ingest sugar, D1 and D2 MSNs compete to resolve a conflict between neural systems signaling gustatory/nutritive value versus those signaling appetite suppression (Sandoval-Rodríguez et al. 2023). Similarly, D1 versus D2 MSNs also control ingestion of cocaine (Lobo et al. 2010), biasing behavior in favor of its rewarding versus aversive properties (Guzman & Ettenberg 2007, O'Neill et al. 2012).

Causal manipulations also provide support for opponent D1/D2 effects on learning. At the synaptic level, stimulating DA bursts promotes growth in concurrently excited striatal D1 neuron spines [mimicking reinforcement of corticostriatal inputs (Yagishita et al. (2014)]. Conversely, causal suppression of DA activity (i.e., DA dips) induces spine growth in correspondingly activated D2 MSNs (Iino et al. 2020). Over the course of behavioral learning, positive and negative DA deflections promote intracellular signaling and synaptic potentiation in D1 and D2 neurons, respectively (Lee et al. 2020, Urakubo et al. 2021). Electrophysiologically, D1 neurons code for reward outcomes, whereas D2 neurons code for no-reward outcomes (Nonomura et al. 2018). After an animal executes a behavior, optogenetic activation of D1 neurons reinforces that behavior, whereas D2 neuron activation suppresses it (Isett et al. 2023, Kravitz et al. 2012, LeBlanc et al. 2018, Nonomura et al. 2018). Conversely, silencing D1 or D2 neurons abolishes learning from positive and negative outcomes, respectively (Hikida et al. 2010, Nishioka et al. 2023). While it is difficult to isolate D1 from D2 neurons in primates, electrophysiological studies in monkeys report equal prevalence of striatal neurons coding for positive and negative action-outcome values, with positive-coding neurons predictive of choice of the coded action, and negative-coding neurons predictive of avoidance (Samejima et al. 2005).

One notable study directly pitted the classical motor activation/suppression (go/no go) account of D1/D2 neurons against the RL account (Yttri & Dudman 2016). The authors employed a closed-loop stimulation protocol that would stimulate D1 or D2 neurons after actions that were executed faster or slower than usual. The classical account would predict that striatal D1 activation would promote faster movements, and D2 activation would promote slower movements. Instead, the authors found that whether the movements were fast or slow, D1 neuron activation would reinforce that speed, and D2 neuron activation would punish that speed (i.e., animals would move faster if D2 neurons were activated after slow movements). These findings provide strong support for a motivational reinforcement impact of D1/D2 opponency rather than go/no go function.

While it is relatively uncontroversial that the BG and DA contribute to reward-oriented behavior, that D2 neurons can support avoidance of actions with the highest cost is more controversial and is thus treated separately here.[2] Importantly, the causal impact of striatal D2 neurons on avoiding cost generalizes across multiple forms of cost. In rodents, activation of D2 neurons suppresses

[2]By avoidance, here I refer to passive avoidance—that is, D2 neurons suppress actions that would produce costly outcomes. This is distinct from active avoidance, whereby animals learn to actively select new instrumental actions to prevent an expected aversive outcome. Indeed, in active avoidance, the lack of expected aversive outcomes produces positive DA RPEs and D1-dependent learning to reinforce the associated actions (Wenzel et al. 2018, Wietzikoski et al. 2011).

the repetition of response errors when seeking reward as well as those that produce overtly aversive outcomes (Danjo et al. 2014, Nishioka et al. 2023). Moreover, endogenous DA dips during negative outcomes are associated with a subsequent increase in striatal D2 neuron activity, which is necessary for the animal's ability to then avoid the corresponding action (Danjo et al. 2014, Nishioka et al. 2023, Nonomura et al. 2018). D2 neurons and their projections along the indirect pathway to globus pallidus are also needed for decisions to avoid physical effort cost (Mingote et al. 2008, Salamone & Correa 2024); this projection is also causally implicated in learning from punishment (Isett et al. 2023). Pharmacological modulation of D2 neuron excitability bidirectionally influences effort costs, with drugs that increase excitability enhancing effort costs and those that suppress D2 neurons reducing effort cost (Farrar et al. 2010, Salamone & Correa 2024). These D2 manipulations specifically modulate the impact of effort costs and do not affect reward sensitivity (Bailey et al. 2020). Striatal D2 neurons also code for the cost of risky decisions, activating when an animal fails to obtain a reward after a risky choice (Zalocusky et al. 2016). Moreover, optogenetically activating these neurons during choices causes normally risk-seeking animals to avoid risky choices and to actively select the safe choice (Zalocusky et al. 2016)—thereby not simply suppressing motor actions writ large but preferentially those with the highest costs. Similar results are obtained by inhibiting DA neurons to induce risk avoidance or by stimulating them to induce risk preference (Stopper et al. 2014). Finally, pharmacological drugs that suppress or activate D2 neurons increase and decrease risk taking, respectively (St. Onge et al. 2010, Zalocusky et al. 2016).

In humans, although it is impossible to causally perturb D1 and D2 populations, many studies have found analogous findings via pharmacological manipulations: Drugs that increase DA amplify the impact of benefits versus costs on learning, with corresponding changes in striatal RPE correlates (Cools 2006; Frank et al. 2004, 2007; Jocham et al. 2011; McCoy et al. 2019; Smittenaar et al. 2012; van der Schaaf et al. 2014; Voon et al. 2010). These RL studies are complemented by very similar DA manipulation effects in studies for which benefits and costs are made explicit with financial gains versus effort costs [including both physical and cognitive effort (Chong et al. 2015, Le Bouc et al. 2016, Pagnier et al. 2024, Westbrook et al. 2020)] and in risky decision making (Rutledge et al. 2015). These effects are also clinically relevant and can account for the impact of DA depletion (e.g., in Parkinson's disease) or D2 blockade (as a treatment for schizophrenia) on avoidance behaviors and, conversely, the impact of DA-promoting medications on risky decision making (Beeler et al. 2012, Frank et al. 2007, Maia & Frank 2017, Voon et al. 2010). Moreover, although they preclude causal interpretations, positron emission tomography (PET) studies measuring striatal D1/D2 receptor availabilities and genetic polymorphism studies have consistently shown that variation in D1 versus D2 function is differentially related to choices based on positive versus negative outcomes (Cox et al. 2015, Doll et al. 2011, Frank 2011, Frank et al. 2007).

A RATIONALE FOR OPPONENCY: THE OPAL* MODEL OF STRIATAL COST-BENEFIT CONTROL

As reviewed above, the evidence for opponent striatal D1/D2 control over cost-benefit learning and decision making is extensive, but it is inconsistent with the simplest classical motor activation versus suppression models. Nevertheless, these findings do not address the question of why the brain might have evolved this opponent system in the first place. Indeed, in computer science, RL models can effectively maximize their cumulative rewards without any notion of opponency. A clue for what is missing in these models is that engineers typically optimize RL hyperparameters for any given application. We recently analyzed RL agent performance in a series of simple learning problems with variable task contexts, including the number of available actions to learn about and the overall reward richness of the environment (i.e., in some tasks rewards were

plentiful, and in others they were sparse). We found that standard RL agents needed quite different parameter settings to perform well across these task environments, and in general they performed better in reward-rich than reward-sparse settings (Collins & Frank 2014, Jaskir & Frank 2023). This task dependence is not an issue for industrial applications, where parameters can be optimized, but we considered whether biological architectures might provide some robustness against this issue, allowing animals to learn effectively across environments. Indeed, animals do not suffer from worse accuracy in reward-sparse than rich settings of the type simulated (Hamid et al. 2015).

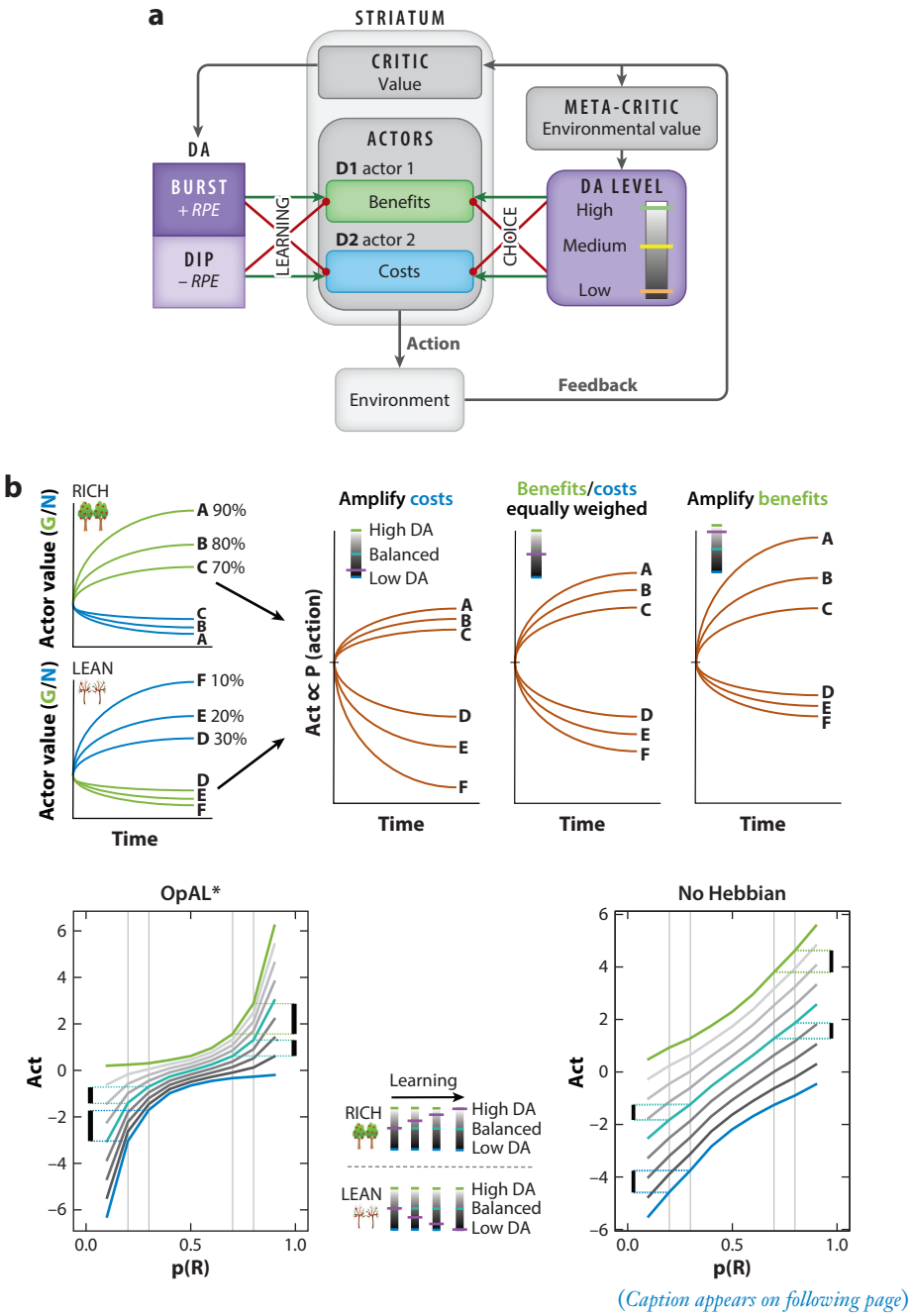
To address this question, let us examine a simplified model of the striatal system called the opponent actor learning model (OpAL, pronounced *opal*, like the gemstone) (Collins & Frank 2014). OpAL extracts core principles from the more detailed neural network models in algorithmic form and is a biological variant of an actor-critic model, often used in RL. The critic learns the expected values of the current state (or state-action pair) and generates RPEs that are used to adjust these expectations. The actor selects an action and then uses the critic-generated RPEs to adjust its policy, that is, its weights for preferring one action or another. The critic is thought to involve regions that target and control midbrain DA, whereas the actor controls behavioral output via its impacts on downstream thalamocortical and brainstem DA.

In OpAL, there are two actors (representing the D1 and D2 populations), which learn in opposite directions from RPEs, and choices are made by combining and weighting them in opponent fashion. Note that such opponency, on its own, is redundant—as the D1 actor increases its weights, the D2 actor would decrease its weights, and vice versa, such that each actor equally discriminates between actions with varying reward probabilities. This symmetry is broken, however, in the biological learning rules adopted by OpAL, which, over the course of learning, distorts what is learned by each actor. In particular, synaptic learning in striatal neurons is activity dependent, influenced jointly by DA and the postsynaptic excitability of the corresponding striatal neuron (Iino et al. 2020, Scott & Frank 2023, Shen et al. 2008, Yagishita et al. 2014). In the neural networks, this activity-dependent learning rule allows learning to occur specifically for those neurons coding the relevant sensory state and action. In OpAL, this Hebbian nonlinearity produces a recursive update rule (i.e., $\Delta D1 \propto D1 \times RPE$), such that D1 actors increase their weights for actions that produce many positive RPEs and, in turn, become more active and hence proportionally more eligible for further learning. Conversely, when negative RPEs are more prevalent, the D2 actors increase their weights and excitability and become more eligible for subsequent learning ($\Delta D2 \propto D2 \times -RPE$). As such, over the course of learning, the D1 and D2 actors come to differentially specialize in discriminating between benefits versus costs of alternative actions (**Figure 1*b***). By expanding the recursive update rule, we showed that each weight update is a function of the entire history of RPEs, amplifying actor weights for actions that had produced sequences of RPEs that are largely consistent in sign (Jaskir & Frank 2023) (in contrast to standard RL models in which weight updates are only proportional to the most recent RPE). This produces convexity in the learned weights that are nonredundant across actors; for example, D1 weights show greater differentiation between actions that yield 90% versus 80% probabilities of reward than they do for 80% versus 70%, and they show little discrimination between small reward probabilities. In contrast, D2 actor weights show strong discrimination between actions with sparse reward probabilities due to the consistent prevalence of negative RPEs (**Figure 1**).

What is the implication of such convexity in D1 and D2 actors? Note that when actors are equally weighted during choice (i.e., intermediate levels of DA), the convexity largely cancels out and the agent is sensitive to differences in both high and low reward probability (Collins & Frank 2014). Critically, however, dynamic DA adjustments can reweight which actor primarily contributes to choice. When in a high-DA state, choices are determined largely by differences in

the D1 actor weights (the benefits), but when in a low-DA state, choices are determined by the D2 actor weights (the costs).

The original OpAL model accounted for many of the effects of exogenous DA manipulations on cost-benefit learning and decision making reviewed above. We recently developed an upgrade to OpAL, called OpAL*, which allows the agent to dynamically modulate its own DA levels



(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

The opponent actor learning model (OpAL*) model of striatal cost-benefit control (Jaskir & Frank 2023) (*a*) D1 and D2 actors contribute to action selection and are subject to reinforcement learning. The critic learns expected values and generates dopamine (DA) reward prediction errors (RPEs). DA bursts potentiate D1 actor weights for the chosen action and depotentiate D2 actor weights. DA dips exert the opposite effect, potentiating D2 actor weights. The meta-critic tracks environmental reward statistics and controls basal striatal DA levels (distinctly from transient RPEs used for learning), differentially weighting D1 versus D2 actor contributions during choice. (*b*, *top*) Evolution of D1 (*green*) and D2 (*orange*) actor weights over trials for various actions with different reward probabilities. Nonlinear Hebbian plasticity enables D1 actors to specialize in discriminating between benefits of options (the difference between 90% and 80% is greater than between 80% and 70%; i.e., actor weights are convex). The benefits of low reward options are deemphasized. Conversely, D2 actors enhance their weights with frequent negative RPEs and come to specialize in discriminating between the costs (here, sparse rewards). The net choice function is subject to DA levels: When they are balanced, benefits and costs are equally weighed, but high/low DA increases discrimination between benefits and costs, respectively. (*b*, *bottom*) The net decision output of the OpAL* actor is convex with respect to reward probability, depending on DA levels during choice. Black bars indicate the difference in actor output for actions with two different reward probabilities under high (*green*), intermediate (*yellow*), and low (*orange*) lines. For high DA, the actor increasingly emphasizes differences in high-probability rewards but de-emphasizes differences at the low end, and vice versa for low DA. Without nonlinear Hebbian plasticity, these curves are parallel, indicating redundancy in the two actors. Figure adapted from Jaskir & Frank (2023) (CC BY 4.0).

(**Figure 1*a***). Intriguingly, the history-dependence and convexity of such effects is reminiscent of work showing that human subjects' moods are sensitive to the momentum of previous RPEs (Eldar et al. 2016). Indeed, recent evidence implies that, even endogenously, striatal DA levels vary within an animal as a function of its environmental context: DA levels ramp up in reward-rich environments, and increasingly so as the frequency of recent rewards increases (Hamid et al. 2015, Mohebi et al. 2019). Such ramps are also prolonged when task demands require instrumental approach actions compared to equivalent Pavlovian conditions, suggesting that DA levels can be adaptively controlled (Goedhoop et al. 2023, Hamid et al. 2021, Lloyd & Dayan 2023, Syed et al. 2015).

We thus augmented OpAL* to include a meta-critic that keeps track of the reward history of the environment and accordingly regulates striatal DA. [Mechanistically, this is thought to involve prefrontal cortical regions that track task states and modulate striatal DA release via inputs to striatal cholinergic neurons (Berke 2018, Mohebi et al. 2019, Stalnaker et al. 2016)]. As such, OpAL* dynamically reweights the contributions of its actors to suit the environmental task context, prioritizing discrimination of either benefits or costs. Large-scale simulations across a wide range of parameters showed that this scheme is useful. In particular, OpAL* was more robust to variations across task environments than various standard RL models, which needed their parameters to be tuned for each task, given different demands on exploration (including models designed to have sophisticated exploration strategies). OpAL* performance advantages were particularly prevalent in ecologically valid, reward-sparse environments with many choice alternatives (Jaskir & Frank 2023). Moreover, these OpAL* advantages depended on three key biological properties for them to manifest: opponency (allowing changes in policies that can recruit one actor or another), nonlinear Hebbian synaptic plasticity rules (allowing each actor to specialize via convexity), and dynamic DA modulation (allowing the agent to reweight which actor is leveraged for choice according to its specialization).

Aside from its normative advantage, OpAL* also accounted for a variety of new empirical data ranging from optogenetics to pharmacological manipulations and behavioral economics (Jaskir & Frank 2023). For example, the model showed how it can adaptively modulate risk taking for potential benefits and that this depends on dynamic DA; this mechanism was sufficient to

account for gambling experiments in which humans taking levodopa (enhancing DA release) increase risk taking in the context of potential gains but not losses (Rutledge et al. 2015). It also accounted for impacts of striatal D2 neuron activity following unsuccessful risky choices in rodents, whereby risk-seeking rodents can be causally impelled to make a safe choice if D2 neurons are optogenetically stimulated during choice (Zalocusky et al. 2016).

Although it is biologically inspired, OpAL* also generated two key predictions that could be compared against existing choice patterns even without DA manipulations. First, when given an offer to make a risky gamble, given identical outcome payoffs, humans are more likely to gamble if this offer is presented in the context of a rich history of gamble outcomes compared to a lean distribution (Frydman & Jin 2022). Economists interpreted their effects by analogy to efficient coding principles, where choices are sensitive to outcomes that are more frequently observed in the environment. According to OpAL*, these efficient coding mechanisms are imparted by convexity specialization within opponent actors, with adaptive DA modulation across rich and lean contexts—facilitating context-dependent reward optimization. The DA dependence of these effects remains a key prediction of OpAL*.

Second, because it eschews value learning within its actors and instead expediently optimizes a policy, OpAL* is also consistent with findings that human choice preferences [and striatal activity (Li & Daw 2011)] are aligned with policy rather than value learners (Palminteri et al. 2015). For example, consider a gain context in which people have to learn to choose between an option that is 75% rewarded and one that is only 25% rewarded, and otherwise produces nothing. In the loss context, they choose between an option that yields a loss 75% of the time versus a loss-avoiding option that still produces a loss 25% of the time (and otherwise nothing). As expected, participants will choose the 75% rewarding option in the gain context and the 25% loss option in the loss context. However, if they are later asked to make transitive preferences across options, they are actually more likely to choose the 25% loss option over the 25% rewarding option, despite the fact that the latter option has higher expected value (Gold et al. 2013, Palminteri et al. 2015). These results were reproduced by OpAL*, as it learns to optimize a policy based on RPEs that had been encountered in corresponding contexts (Jaskir & Frank 2023). This same OpAL* mechanism is consistent with rodent findings whereby striatal DA deflections (whether naturally occurring or causally stimulated) induce a change in policy rather than value learning (Coddington et al. 2023). While these behaviors challenge the normative principle (participants should ultimately be able to prefer options that have higher expected value), we consider these behaviors to reflect a byproduct of a normative mechanism that expediently learns an optimized policy for a given context. From this perspective, it would still be useful to learn expected values (as OpAL* does within its critic), but simulations showed that it takes longer for such expected values to converge than for the optimal OpAL* policy to stabilize.

OpAL* is not the only normative proposal for BG and DA contributions to choice. A recent modeling study suggested that using novelty to boost DA signaling (rather than environmental reward richness) can be useful to optimize the exploration/exploitation dilemma (Wang et al. 2024). These models share many of the same core principles but lack the Hebbian component and instead have different nonlinearities that render the opponent actors nonredundant.[3] Nevertheless, this model is also able to account for some DA manipulation data, including impacts on risky choice.

[3]In this model, positive and negative RPEs are asymmetrically encoded in D1 versus D2 actors, but weight updates are not additionally activity dependent. In principle the two nonlinearities can be combined at the cost of a more complex model [indeed, they are both features of the neural network models described earlier; for a discussion, see Collins & Frank (2014)]. Thus, we do not challenge asymmetries in sensitivities to positive and negative RPEs; we simply highlight the additional role of activity dependence, which is sufficient for OpAL

The value of computational modeling is being explicit in the formulations, leading to testable predictions that can discriminate between candidate models and ultimately improve upon them. In this spirit, I articulate one key aspect that distinguishes OpAL from these and other models. In particular, recall that the Hebbian nonlinearity induces convexity in the learned actor weights. This nonlinearity is adaptive, as noted above, but it is also needed to account for a wide variety of data across species (Collins & Frank 2014, Jaskir & Frank 2023). For example, DA medications make humans more sensitive to discriminating between reward-rich choice options (e.g., 80% versus 70% and 60% reward probability) but worse at discriminating between lean choice options (e.g., 20% versus 30% and 40%). Conversely, low DA levels actually enhance discrimination between low reward values or costs (Cockburn et al. 2014; Frank et al. 2004, 2007; McCoy et al. 2019). These patterns (which have been replicated more than 10 times) are direct consequences of the OpAL convexity and are observed even if DA manipulations are administered after learning—suggesting that they are in part mediated by DA effects on motivated choice (reweighting of actor contributions in OpAL*) rather than only asymmetries in learning (Shiner et al. 2012, Smittenaar et al. 2012). In contrast, actor weights in other opponency models are concave rather than convex and thus predict the opposite pattern than that observed empirically (i.e., where high DA levels favor discrimination between low instead of high reward probabilities, and vice versa) (Mikhael & Bogacz 2016).

While the above studies focused on learning (and later expression thereof), similar patterns are seen in DA manipulation studies on cost-benefit choices about physical and cognitive effort. Here the OpAL convexity predicts that higher DA levels should induce steeper effects on the psychometric choice function when benefits outweigh the costs but blunted effects when costs outweigh benefits (**Figure 2**). This pattern was observed in humans with higher levels of striatal DA release (as assessed with ^{18}F-DOPA PET imaging), and it was also causally induced in low-DA subjects after taking a dose of the stimulant methylphenidate (Westbrook et al. 2020). This finding is directly analogous to that seen in rodents reviewed above, where D1 and D2 inactivation impact choices on opposite sides of the psychometric function (Bolkan et al. (2022) (**Figure 2**). Moreover, while this study focused on decisions about cognitive effort—recasting the impact of stimulants as impacting motivational cognitive factors rather than ability, per se—similar cost-benefit effects of DA manipulations were recently observed in a physical effort in Parkinson's disease (Pagnier et al. 2024).

A particularly conspicuous set of studies suggestive of nonlinear Hebbian plasticity and convexity come from rodents administered D2-blocking drugs. Despite the nonlinearity, in OpAL*, actor weights do not grow without bound, and stabilize due to adjustments in critic expectations, leading to smaller RPEs (Collins & Frank 2014, Jaskir & Frank 2023). However, this stabilizing effect should be disrupted under DA manipulations. Indeed, when D2-blocking drugs are administered in the same context over days, animals become progressively slower to move, but only in the context in which the D2 blocker is applied. This context-dependent catalepsy sensitization was modeled in one of our early neural network models, wherein D2 blockade enhances excitability of D2 neurons, which then undergo Hebbian potentiation, making them yet more excitable and in turn more eligible for further aberrant learning, and so on (Wiecki et al. 2009). Later work confirmed that these drugs potentiate striatal D2 neurons and that this mechanism can similarly induce progressive impairments in motor skill learning, which then persist after drug washout

to generate actors that specialize on benefits and costs. Moreover, while Wang et al. (2024) highlighted an advantage of their model over OpAL* in some task contexts, our subsequent simulations revealed that this was artifactual and that both models performed favorably, with OpAL* still showing some advantages (J.T. Hewson, A. Jaskir & M.J. Frank, unpublished manuscript).

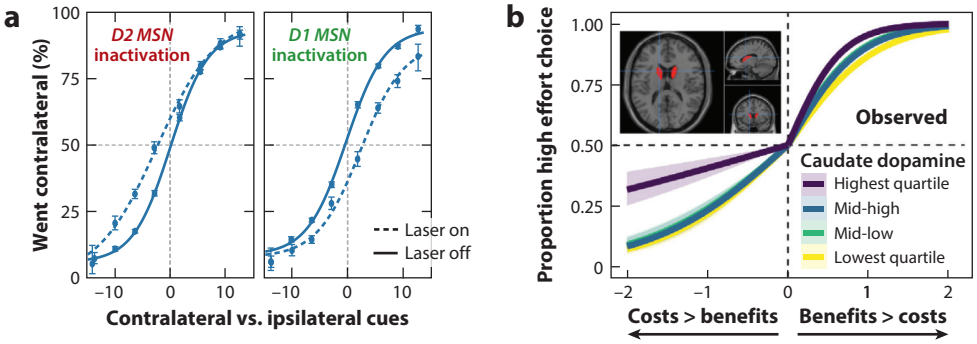


Figure 2

Cross-species evidence for opponent striatal DA modulation of benefit and cost decision making. (*a*) Mice performing a decision-making task requiring them to accumulate the number of cues observed on contralateral versus ipsilateral sides of space. When engaged, animals were highly sensitive to the number of cues. Striatal D1 inactivation in this condition selectively reduced the impact of positive evidence for contralateral cues, whereas D2 inactivation reduced the impact of negative evidence against contralateral cues. Dashed lines reflect performance after inactivation. (*b*) Analogous findings in humans performing a cognitive effort task for varying amounts of benefits and costs. Individual differences in striatal DA were related to the increased slope of the psychometric function to the right of the indifference point (i.e., when benefits outweigh costs) but render this curve more shallow on the left (when costs outweigh benefits). These patterns were also induced by stimulants that increase striatal DA release and accorded with simulations from the OpAL model (not shown). Abbreviations: DA, dopamine; MSN, medium spiny neuron; OpAL, opponent actor learning model. Panel *a* adapted with permission from Bolkan et al. (2022), and panel *b* adapted with permission from Westbrook et al. (2020).

(Beeler et al. 2012, Cheung et al. 2023). Again, these effects required the nonlinear Hebbian plasticity rule for them to manifest in both the neural network and OpAL (Beeler et al. 2012, Collins & Frank 2014). Moreover, these aberrant learning effects can be rescued by coadministering adenosine blockers that prevent long-term potentiation in D2 neurons (Beeler et al. 2012). In sum, apparent costs of motor action become amplified over experience under DA denervation or D2 blockade. While these are pathological effects (suggesting that some aspects of parkinsonism reflect an aberrant learning process), they can nevertheless be considered a by-product of a normative process when DA levels are endogenously regulated in a healthy brain.

A MOSAIC OF COST-BENEFIT CONTROL ACROSS CORTICOSTRIATAL CIRCUITS

The above discussion might make it seem as if there is a single corticostriatal circuit with two knobs controlling the influences of benefits and costs for each action. However, while there is a canonical circuit linking frontal cortex with the BG and thalamus, this motif is repeated across virtually all of frontal cortex: from motor to premotor cortex to various subdivisions in prefrontal cortex, each interconnected with its own BG-thalamic loop, which also interact hierarchically (Badre & Frank 2012, Frank 2011, Graybiel 2008, Haber 2016, Westbrook et al. 2021). Indeed, the cost-benefit framework articulated above can apply within any given corticostriatal circuit at different levels of abstraction (e.g., What are the benefits and costs of going left versus right, of going to one cafe or another, of performing one cognitive action or another, or of applying one control strategy or another?). At a larger scale, this same cost-benefit mechanism may occur across circuits (e.g., When should we employ hierarchical planning requiring nested coordination across multiple corticostriatal loops versus lay low and let our more primitive motor RL system control

behavior?). Models of this higher-level selection process adopt a mixture-of-experts scheme, where the costs and benefits of computations at different levels of abstraction are compared and dynamically recruited as needed, and can be implemented in hierarchically nested corticostriatal circuits (Frank & Badre 2012). Supporting this scheme, when humans performed a task with a hierarchical structure, RPEs detected in functional MRI (fMRI) were not globally uniform across striatum but were amplified in caudate subregions interconnected with prefrontal regions at the same rostrocaudal level tuned to hierarchical task structure (Badre & Frank 2012). Moreover, when the task no longer required hierarchical processing, striatal negative RPEs were associated with participants' ability to inactivate that cortical region and learn via alternate strategies.

Of course, fMRI is an indirect measure of neural activity. Recent rodent studies have recorded activity in dopaminergic axons or release within multiple striatal regions, often focusing on ventral, dorsomedial, and dorsolateral striatum—functional subdivisions that have homologs in primates. These studies have shown both consistency and heterogeneity in DA signals across regions (Hamid et al. 2021, Mohebi et al. 2024, Parker et al. 2016, Tsutsui-Kimura et al. 2020), where all exhibited RPE-like transients, but tailored to the computations of the underlying region. For example, DA RPE signals in ventral striatum reflect deviations in stimulus-reward expectations, whereas those in dorsomedial striatum (DMS) were tied to specific instrumental actions (Parker et al. 2016), converging with coding of the underlying striatal neurons within these regions (Ito & Doya 2015). Other studies indicate a gradient of DA RPE dynamics across these regions, with the slow dynamics in ventral striatum sensitive to longer-term horizons of reward, and the faster dynamics within dorsolateral striatum sensitive to more immediate rewards (Mohebi et al. 2024). Finally, DA neurons themselves exhibit heterogeneity in their scaling of RPEs, resembling a distributional code (Dabney et al. 2020), which could conceivably differentially impact distinct striatal subregions.

How can the brain learn the benefits and costs of using a given striatal region to control behavior? To address this question, we recently studied DA dynamics within subregions of dorsal striatum (Hamid et al. 2021). Mice had to learn in separate instrumental and Pavlovian task conditions, which were reversed across blocks of trials. In both conditions, they experienced a series of sensory cues indicating progress to reward. In the instrumental condition, mice had to run on a wheel for the cues to advance and to ultimately obtain reward. In the Pavlovian condition, the cues advanced irrespective of the animal's behavior. The amount of running needed (or the time between cues) was drawn from a uniform distribution, requiring evidence accumulation across cues to infer whether the animal was in control of the cues (and needed to run). Notably, DA signals ramped as the animal progressed to reward, but these ramps were heterogeneous across striatal subregions and task conditions. In the instrumental condition, increasing DA ramps were observed in DMS, with focal subregions within DMS showing DA ramps tuned to different instrumental contingencies (distance to run). Moreover, in the Pavlovian condition, dorsomedial DA ramps were actually negative (i.e., DA levels declined over the course of the trial), despite the fact that mice continued to anticipate reward with each cue transition (as indicated by their licking). These results were modeled with a hierarchical mixture-of-experts framework in which subexperts within DMS represent action-outcome contingencies, and DMS as a whole accumulates evidence that it is in control when actions are tied to state transitions—and more evidence that is not in control in the Pavlovian condition. Thus, DA ramps in DMS seem to reflect the value of its underlying computations. When DA levels rise, they presumably engage the local D1 neuron subexperts to control behavior (subject to their own evidence accumulation as per the above discussion), and when local DA levels decline, other regions take over.

A yet more striking result from this study was seen when mice obtained their reward. While there was a massive increase in DA levels across the striatum, these levels were not globally synchronous but instead propagated in spatiotemporal wave-like patterns. Reward-induced waves

reached DMS first in the instrumental condition but last in the Pavlovian condition; these dynamics reversed when the task contingencies switched (Hamid et al. 2021). These wave-like DA dynamics were predictive of the animal's running more or less on subsequent trials—and were themselves predicted by the preceding ramp dynamics signaling evidence of control. Thus, DA ramps and wave dynamics may serve as a credit assignment mechanism to reinforce DMS when the benefits of action outweigh the cost. Finally, this study also observed rapid DA transients at each cue transition as the animal progressed through the trial. These transients exhibited RPE-like properties that were separate from ramping signals, and they could be used to infer agency of the underlying subregion. In sum, this study showed how disparate DA dynamics (transients, ramps, and waves) could work in concert to support inference about agency and credit assignment therein. Thus, the corticostriatal hierarchy may facilitate an iterative evaluation of lower-level predictions to guide action selection and learning at the appropriate level of abstraction.

Humans may show similar cost-benefit trade-offs for recruiting different corticostriatal circuits. When people switch between tasks that involve controllable or uncontrollable state transitions, striatal and medial prefrontal prediction error signals distinguish between the corresponding actor and spectator models that govern these statistics (Ligneul et al. 2022) and thus may be used for credit assignment. Striatal DA contributions to learning of benefits versus costs are also enhanced by agency [free versus forced choices (Cockburn et al. 2014)]. Moreover, even in simple instrumental learning tasks, participants adopt cognitive processing and recruit working memory (Collins & Frank 2012). This too can be construed as a cost-benefit problem: Working memory is rapid and flexible but computationally costly; indeed, people's decisions about performing working memory operations are themselves subject to striatal DA modulation of benefits and costs (Westbrook et al. 2021, 2020). When a learning task is too demanding on working memory, participants show slowed acquisition of task contingencies but enhanced RPE signaling and neural RL signals (Collins & Frank 2018), which are in turn predictive of better long-term retention of stimulus-response associations (Rac-Lubashevsky et al. 2023). Finally, even within the domain of working memory, one can adaptively switch between strategies that support precise storage and recollection of individual items versus those that can chunk multiple items into a single merged representation to improve memory efficiency at the cost of precision when the task is more difficult (Nassar et al. 2018). Such adaptive chunking can be learned via dopaminergic RL signals that optimize the benefits and costs of alternative gating policies in corticostriatal network models (Soni & Frank 2024). Conversely, failures in adaptive gating can lead to deficits in meta-control that may induce rumination and worry (Hitchcock & Frank 2024).

DISCUSSION AND OUTSTANDING ISSUES

The segregation of direct and indirect pathways has long been controversial. Most recently, it has been pointed out that while D2 neurons project solely to globus pallidus (the classical indirect pathway), D1 neurons project to both substantia nigra (classical direct pathway) and globus pallidus (Lévesque & Parent 2005). While these anatomical data suggest some need of revision, they do not necessarily demand wholesale reevaluation of opponent dichotomies, especially given the large body of functional evidence. Indeed, because of the additional inhibitory synapse and associated delay, such connectivity may simply imply that BG output is sensitive to the temporal derivative of striatal D1 neuron activity rather than its raw activity, facilitating rapid sequencing of actions (Frank 2006). Moreover, there are certainly functional studies that do not conform naturally to the valenced opponency account articulated above, for example, in accumbens (Zachry et al. 2024). While a mixture-of-experts framework might imply that other

striatal subregions perform these computations depending on the task at hand, future theoretical and empirical studies should incorporate and reconcile challenging as well as supportive data.

I have emphasized how modulation of striatal DA signals can impact cost-benefit decision making, including about risk. It is notable that DA neurons themselves may also provide a distributional code of potential reward values (Dabney et al. 2020); such a code could feasibly be used for adaptive risk taking, but it remains to be studied how it might interact with striatal D1/D2 opponency.

As noted at the outset, it is also controversial whether dopaminergic transients are themselves always valenced or whether they serve as generalized prediction error signals. While I emphasized that the answer may depend on the target region, even striatal DA transients might not always induce opponent plasticity mechanisms. Indeed, plasticity is gated by additional factors beyond DA, such as pauses in cholinergic signaling, which can adapt the rate of learning across striatal populations (Franklin & Frank 2015). However, cholinergic signals may themselves gate DA release and ramping (Berke 2018), and we are far from a unified model of these interactions. Even less is known about the mechanisms that drive DA waves (Hamid et al. 2021).

A notable recent finding is that control over DA is itself subject to D1 and D2 opponency within striosomal compartments of the striatum, complementing the direct and indirect pathways for control of action (Lazaridis et al. 2024). This finding may imply a fractal structure, whereby opponency shapes valuation and choice simultaneously, or perhaps via ascending spiraling loops (Haber 2016). At the computational level, it has been argued that DA levels may be regulated when Pavlovian tendencies interfere with instrumental goals (Lloyd & Dayan 2023), and it is plausible that such opponency allows the costs and benefits of DA release to be evaluated in such circumstances.

The striatum is not the only region of the brain involved in cost-benefit decision making. Indeed, there is a parallel literature on limbic regions within prefrontal cortex, such as anterior cingulate, ventromedial, and orbitofrontal cortex, that undoubtedly participate in such computations (Klein-Flügge et al. 2016, Shenhav et al. 2013), and may provide that information to striatum for adaptive behavior without requiring new learning. More broadly, a major challenge to the field is to define what constitutes a benefit and cost in the first place (Juechems & Summerfield 2019), which requires understanding across multiple scales from mechanism (e.g., inputs to DA signals) to computation and even philosophy. In realistic scenarios, the benefits and costs of alternative choices are multidimensional (Hall et al. 2024) (hence requiring integration), and more work is needed to study how these computations converge into DA and striatal systems for learning and choice.

DISCLOSURE STATEMENT

M.J.F. has received consulting fees from F. Hoffman-La Roche, Gilgamesh Pharmaceuticals, and Boehringer-Ingelheim.

ACKNOWLEDGMENTS

I thank all of my current and former mentors, mentees, and collaborators for shaping and refining many of these ideas. I thank Roshan Cools, Alana Jaskir, Arif Hamid, Guillaume Pagnier, and an anonymous reviewer for helpful comments on an earlier version of this manuscript, and for each making seminal contributions of their own on these topics.

LITERATURE CITED

Abe K, Kambe Y, Majima K, Hu Z, Ohtake M, et al. 2024. Functional diversity of dopamine axons in prefrontal cortex during classical conditioning. *eLife* 12:RP91136

Albin RL, Young AB, Penney JB. 1989. The functional anatomy of basal ganglia disorders. *Trends Neurosci.* 12(10):366–75

Ashby FG, Ennis JM, Spiering BJ. 2007. A neurobiological theory of automaticity in perceptual categorization. *Psychol. Rev.* 114(3):632–56

Badre D, Frank MJ. 2012. Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: evidence from fMRI. *Cereb. Cortex* 22(3):527–36

Bailey MR, Chun E, Schipani E, Balsam PD, Simpson EH. 2020. Dissociating the effects of dopamine D2 receptors on effort-based versus value-based decision making using a novel behavioral approach. *Behav. Neurosci.* 134(2):101–18

Bayer HM, Lau B, Glimcher PW. 2007. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J. Neurophysiol.* 98(3):1428–39

Beeler JA, Frank MJ, McDaid J, Alexander E, Turkson S, et al. 2012. A role for dopamine-mediated learning in the pathophysiology and treatment of Parkinson's disease. *Cell Rep.* 2(6):1747–61

Berke JD. 2018. What does dopamine mean? *Nat. Neurosci.* 21(6):787–93

Berridge KC. 2006. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* 191(3):391–431

Bogacz R, Gurney K. 2007. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput.* 19(2):442–77

Bolkan SS, Stone IR, Pinto L, Ashwood ZC, Iravedra Garcia JM, et al. 2022. Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nat. Neurosci.* 25(3):345–57

Brown R, Marsden C. 1988. Internal versus external cures and the control of attention in Parkinson's disease. *Brain* 111(2):323–45

Burke DA, Rotstein HG, Alvarez VA. 2017. Striatal local circuitry: a new framework for lateral inhibition. *Neuron* 96(2):267–84

Busemeyer JR, Gluth S, Rieskamp J, Turner BM. 2019. Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends Cogn. Sci.* 23(3):251–63

Calderon CB, Verguts T, Frank MJ. 2022. Thunderstruck: the ACDC model of flexible sequences and rhythms in recurrent neural circuits. *PLOS Comput. Biol.* 18(2):e1009854

Cheung TH, Ding Y, Zhuang X, Kang UJ. 2023. Learning critically drives parkinsonian motor deficits through imbalanced striatal pathway recruitment. *PNAS* 120(12):e2213093120

Chong TT-J, Bonnelle V, Manohar S, Veromann K-R, Muhammed K, et al. 2015. Dopamine enhances willingness to exert effort for reward in Parkinson's disease. *Cortex* 69:40–46

Cisek P, Kalaska JF. 2005. Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45(5):801–14

Cockburn J, Collins AG, Frank MJ. 2014. A reinforcement learning mechanism responsible for the valuation of free choice. *Neuron* 83(3):551–57

Coddington LT, Lindo SE, Dudman JT. 2023. Mesolimbic dopamine adapts the rate of learning from action. *Nature* 614(7947):294–302

Collins A, Frank MJ. 2012. How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* 35(7):1024–35

Collins A, Frank MJ. 2014. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* 121(3):337–66

Collins A, Frank MJ. 2018. Within and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *PNAS* 115:2502–7

Collins AL, Greenfield VY, Bye JK, Linker KE, Wang AS, Wassum KM. 2016. Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Sci. Rep.* 6(1):20231

Cools R. 2006. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neurosci. Biobehav. Rev.* 30(1):1–23

Corbit LH, Balleine BW. 2011. The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *J. Neurosci.* 31(33):11786–94

Cox SM, Frank MJ, Larcher K, Fellows LK, Clark CA, et al. 2015. Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *NeuroImage* 109:95–101

Cui G, Jun SB, Jin X, Pham MD, Vogel SS, et al. 2013. Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* 494(7436):238–42

Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, et al. 2020. A distributional code for value in dopamine-based reinforcement learning. *Nature* 577(7792):671–75

Danjo T, Yoshimi K, Funabiki K, Yawata S, Nakanishi S. 2014. Aversive behavior induced by optogenetic inactivation of ventral tegmental area dopamine neurons is mediated by dopamine D2 receptors in the nucleus accumbens. *PNAS* 111(17):6455–60

de Jong JW, Afjei SA, Pollak Dorocic I, Peck JR, Liu C, et al. 2019. A neural circuit mechanism for encoding aversive stimuli in the mesolimbic dopamine system. *Neuron* 101(1):133–51.e7

DePasquale B, Brody CD, Pillow JW. 2024. Neural population dynamics underlying evidence accumulation in multiple rat brain regions. bioRxiv 2021.10.28.465122. **https://doi.org/10.1101/2021.10.28.465122**

Doi T, Fan Y, Gold JI, Ding L. 2020. The caudate nucleus contributes causally to decisions that balance reward and uncertain visual information. *eLife* 9:e56694

Doll BB, Hutchison KE, Frank MJ. 2011. Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *J. Neurosci.* 31(16):6188–98

Dunovan K, Vich C, Clapp M, Verstynen T, Rubin J. 2019. Reward-driven changes in striatal pathway competition shape evidence evaluation in decision-making. *PLOS Comput. Biol.* 15(5):e1006998

Eldar E, Rutledge RB, Dolan RJ, Niv Y. 2016. Mood as representation of momentum. *Trends Cogn. Sci.* 20(1):15–24

Farrar A, Segovia K, Randall P, Nunes E, Collins L, et al. 2010. Nucleus accumbens and effort-related functions: behavioral and neural markers of the interactions between adenosine A2A and dopamine D2 receptors. *Neuroscience* 166(4):1056–67

Farrell K, Lak A, Saleem AB. 2022. Midbrain dopamine neurons signal phasic and ramping reward prediction error during goal-directed navigation. *Cell Rep.* 41(2):111470

Ferro D, Cash-Padgett T, Wang MZ, Hayden BY, Moreno-Bote R. 2024. Gaze-centered gating, reactivation, and reevaluation of economic value in orbitofrontal cortex. *Nat. Commun.* 15(1):6163

Frank MJ. 2005. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *J. Cogn. Neurosci.* 17(1):51–72

Frank MJ. 2006. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw.* 19(8):1120–36

Frank MJ. 2011. Computational models of motivated action selection in corticostriatal circuits. *Curr. Opin. Neurobiol.* 21(3):381–86

Frank MJ, Badre D. 2012. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. *Cereb. Cortex* 22(3):509–26

Frank MJ, Claus ED. 2006. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol. Rev.* 113(2):300–26

Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, et al. 2015. fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. *J. Neurosci.* 35(2):485–94

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. 2007. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *PNAS* 104(41):16311–16

Frank MJ, Samanta J, Moustafa AA, Sherman SJ. 2007. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science* 318(5854):1309–12

Frank MJ, Seeberger LC, O'Reilly RC. 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306(5703):1940–43

Franklin NT, Frank MJ. 2015. A cholinergic feedback circuit to regulate striatal population uncertainty and optimize reinforcement learning. *eLife* 4:e12029

Frydman C, Jin LJ. 2022. Efficient coding and risky choice. *Q. J. Econ.* 137(1):161–213

Gerfen CR, Surmeier DJ. 2011. Modulation of striatal projection systems by dopamine. *Annu. Rev. Neurosci.* 34:441–66

Glaser JI, Perich MG, Ramkumar P, Miller LE, Kording KP. 2018. Population coding of conditional probability distributions in dorsal premotor cortex. *Nat. Commun.* 9(1):1788

Gluth S, Spektor MS, Rieskamp J. 2018. Value-based attentional capture affects multi-alternative decision making. *eLife* 7:e39659

Goedhoop J, Arbab T, Willuhn I. 2023. Anticipation of appetitive operant action induces sustained dopamine release in the nucleus accumbens. *J. Neurosci.* 43(21):3922–32

Gold JM, Strauss GP, Waltz JA, Robinson BM, Brown JK, Frank MJ. 2013. Negative symptoms of schizophrenia are associated with abnormal effort-cost computations. *Biol. Psychiatry* 74(2):130–36

Graybiel AM. 2008. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31:359–87

Guzman D, Ettenberg A. 2007. Runway self-administration of intracerebroventricular cocaine: evidence of mixed positive and negative drug actions. *Behav. Pharmacol.* 18(1):53–60

Haber SN. 2016. Corticostriatal circuitry. *Dialogues Clin. Neurosci.* 18(1):7–21

Hall AF, Browning M, Huys QJ. 2024. The computational structure of consummatory anhedonia. *Trends Cogn. Sci.* 28(6):541–53

Hamid AA, Frank MJ, Moore CI. 2021. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* 184(10):2733–49.e16

Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, et al. 2015. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* 19:117–26

Hart AS, Rutledge RB, Glimcher PW, Phillips PEM. 2014. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J. Neurosci.* 34(3):698–704

Hikida T, Kimura K, Wada N, Funabiki K, Nakanishi S. 2010. Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron* 66(6):896–907

Hitchcock PF, Frank MJ. 2024. From tripping and falling to ruminating and worrying: a meta-control account of repetitive negative thinking. *Curr. Opin. Behav. Sci.* 56:101356

Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM. 2013. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* 500(7464):575–79

Iino Y, Sawada T, Yamaguchi K, Tajiri M, Ishii S, et al. 2020. Dopamine D2 receptors in discrimination learning and spine enlargement. *Nature* 579(7800):555–60

Isett BR, Nguyen KP, Schwenk JC, Yurek JR, Snyder CN, et al. 2023. The indirect pathway of the basal ganglia promotes transient punishment but not motor suppression. *Neuron* 111(14):2218–31.e4

Ito M, Doya K. 2015. Distinct neural representation in the dorsolateral, dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks. *J. Neurosci.* 35(8):3499–514

Jaskir A, Frank MJ. 2023. On the normative advantages of dopamine and striatal opponency for learning and choice. *eLife* 12:e85107

Jocham G, Klein TA, Ullsperger M. 2011. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J. Neurosci.* 31(5):1606–13

Juechems K, Summerfield C. 2019. Where does value come from? *Trends Cogn. Sci.* 23(10):836–50

Klein-Flügge MC, Kennerley SW, Friston K, Bestmann S. 2016. Neural signatures of value comparison in human cingulate cortex during decisions requiring an effort-reward trade-off. *J. Neurosci.* 36(39):10002–15

Kool W, Botvinick M. 2014. A labor/leisure tradeoff in cognitive control. *J. Exp. Psychol. Gen.* 143(1):131–41

Kravitz AV, Tye LD, Kreitzer AC. 2012. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat. Neurosci.* 15(6):816–18

Lazaridis I, Crittenden JR, Ahn G, Hirokane K, Yoshida T, et al. 2024. Striosomes target nigral dopamine-containing neurons via direct-D1 and indirect-D2 pathways paralleling classic direct-indirect basal ganglia systems. bioRxiv 2024.06.01.596922. **https://doi.org/10.1101/2024.06.01.596922**

Le Bouc R, Rigoux L, Schmidt L, Degos B, Welter M-L, et al. 2016. Computational dissection of dopamine motor and motivational functions in humans. *J. Neurosci.* 36(25):6623–33

LeBlanc KH, London TD, Szczot I, Bocarsly ME, Friend DM, et al. 2018. Striatopallidal neurons control avoidance behavior in exploratory tasks. *Mol. Psychiatry* 25(2):491–505

Lee RS, Sagiv Y, Engelhard B, Witten IB, Daw ND. 2024. A feature-specific prediction error model explains dopaminergic heterogeneity. *Nat. Neurosci.* 27:1574–86

Lee SJ, Lodder B, Chen Y, Patriarchi T, Tian L, Sabatini BL. 2020. Cell-type-specific asynchronous modulation of PKA by dopamine in learning. *Nature* 590(7846):451–56

Leotti LA, Delgado MR. 2011. The inherent reward of choice. *Psychol. Sci.* 22(10):1310–18

Lévesque M, Parent A. 2005. The striatofugal fiber system in primates: a reevaluation of its organization based on single-axon tracing studies. *PNAS* 102(33):11888–93

Li J, Daw ND. 2011. Signals in human striatum are appropriate for policy update rather than value prediction. *J. Neurosci.* 31(14):5504–11

Ligneul R, Mainen ZF, Ly V, Cools R. 2022. Stress-sensitive inference of task controllability. *Nat. Hum. Behav.* 6(6):812–22

Lloyd K, Dayan P. 2023. Reframing dopamine: a controlled controller at the limbic-motor interface. *PLOS Comput. Biol.* 19(10):e1011569

Lobo MK, Covington HE, Chaudhury D, Friedman AK, Sun H, et al. 2010. Cell type–specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* 330(6002):385–90

Lutas A, Kucukdereli H, Alturkistani O, Carty C, Sugden AU, et al. 2019. State-specific gating of salient cues by midbrain dopaminergic input to basal amygdala. *Nat. Neurosci.* 22(11):1820–33

Maia TV, Frank MJ. 2017. An integrative perspective on the role of dopamine in schizophrenia. *Biol. Psychiatry* 81(1):52–66

McCoy B, Jahfari S, Engels G, Knapen T, Theeuwes J. 2019. Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. *Brain* 142(11):3605–20

Mikhael JG, Bogacz R. 2016. Learning reward uncertainty in the basal ganglia. *PLOS Comput. Biol.* 12(9):e1005062

Mikhael JG, Kim HR, Uchida N, Gershman SJ. 2022. The role of state uncertainty in the dynamics of dopamine. *Curr. Biol.* 32(5):1077–87.e9

Mingote S, Font L, Farrar AM, Vontell R, Worden LT, et al. 2008. Nucleus accumbens adenosine A2A receptors regulate exertion of effort by acting on the ventral striatopallidal pathway. *J. Neurosci.* 28(36):9037–46

Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, et al. 2019. Dissociable dopamine dynamics for learning and motivation. *Nature* 570(7759):65–70

Mohebi A, Wei W, Pelattini L, Kim K, Berke JD. 2024. Dopamine transients follow a striatal gradient of reward time horizons. *Nat. Neurosci.* 27(4):737–46

Montague P, Dayan P, Sejnowski T. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16(5):1936–47

Nassar MR, Helmers JC, Frank MJ. 2018. Chunking as a rational strategy for lossy data compression in visual working memory. *Psychol. Rev.* 125(4):486–511

Nishioka T, Attachaipanich S, Hamaguchi K, Lazarus M, de Kerchove d'Exaerde A, et al. 2023. Error-related signaling in nucleus accumbens D2 receptor-expressing neurons guides inhibition-based choice behavior in mice. *Nat. Commun.* 14(1):2284

Nonomura S, Nishizawa K, Sakai Y, Kawaguchi Y, Kato S, et al. 2018. Monitoring and updating of action selection for goal-directed behavior through the striatal direct and indirect pathways. *Neuron* 99(6):1302–14.e5

O'Neill B, Tilley MR, Gu HH. 2012. Cocaine produces conditioned place aversion in mice with a cocaine-insensitive dopamine transporter. *Genes Brain Behav.* 12(1):34–38

Pagnier GJ, Asaad WF, Frank MJ. 2024. Double dissociation of dopamine and subthalamic nucleus stimulation on effortful cost/benefit decision making. *Curr. Biol.* 34(3):655–60.e3

Palminteri S, Khamassi M, Joffily M, Coricelli G. 2015. Contextual modulation of value signals in reward and punishment learning. *Nat. Commun.* 6(1):8096

Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, et al. 2016. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* 19(6):845–54

Pedersen ML, Frank MJ. 2020. Simultaneous hierarchical Bayesian parameter estimation for reinforcement learning and drift diffusion models: a tutorial and links to neural data. *Comput. Brain Behav.* 3(4):458–71

Rac-Lubashevsky R, Cremer A, Collins A, Frank MJ. 2023. Neural index of reinforcement learning predicts improved stimulus-response retention under high working memory load. *J. Neurosci.* 43:3131–43

Ratcliff R, Frank MJ. 2012. Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Comput.* 24(5):1186–229

Rich EL, Wallis JD. 2016. Decoding subjective decisions from orbitofrontal cortex. *Nat. Neurosci.* 19(7):973–80

Roitman JD, Shadlen MN. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* 22(21):9475–89

Rutledge RB, Skandali N, Dayan P, Dolan RJ. 2015. Dopaminergic modulation of decision making and subjective well-being. *J. Neurosci.* 35(27):9811–22

Salamone JD, Correa M. 2024. The neurobiology of activational aspects of motivation: exertion of effort, effort-based decision making, and the role of dopamine. *Annu. Rev. Psychol.* 75:1–32

Samejima K, Ueda Y, Doya K, Kimura M. 2005. Representation of action-specific reward values in the striatum. *Science* 310(5752):1337–40

Sandoval-Rodríguez R, Parra-Reyes JA, Han W, Rueda-Orozco PE, Perez IO, et al. 2023. D1 and D2 neurons in the nucleus accumbens enable positive and negative control over sugar intake in mice. *Cell Rep.* 42(3):112190

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science* 275(5306):1593–99

Scott DN, Frank MJ. 2023. Adaptive control of synaptic plasticity integrates micro- and macroscopic network function. *Neuropsychopharmacology* 48(1):121–44

Sharpe MJ, Chang CY, Liu MA, Batchelor HM, et al. 2017. Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nat. Neurosci.* 20(5):735–42

Shen W, Flajolet M, Greengard P, Surmeier DJ. 2008. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321(5890):848–51

Shenhav A, Botvinick M, Cohen J. 2013. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79(2):217–40

Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, et al. 2012. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. *Brain* 135(6):1871–83

Smittenaar P, Chase HW, Aarts E, Nusselein B, Bloem BR, Cools R. 2012. Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection—learning or performance? *Eur. J. Neurosci.* 35(7):1144–51

Soni A, Frank MJ. 2024. Adaptive chunking improves effective working memory capacity in a prefrontal cortex and basal ganglia circuit. *eLife* 13:RP97894

St. Onge JR, Chiu YC, Floresco SB. 2010. Differential effects of dopaminergic manipulations on risky choice. *Psychopharmacology* 211(2):209–21

St. Onge JR, Floresco SB. 2009. Dopaminergic modulation of risk-based decision making. *Neuropsychopharmacology* 34(3):681–97

Stalnaker TA, Berg B, Aujla N, Schoenbaum G. 2016. Cholinergic interneurons use orbitofrontal input to track beliefs about current state. *J. Neurosci.* 36(23):6242–57

Stauffer WR, Lak A, Yang A, Borel M, Paulsen O, et al. 2016. Dopamine neuron-specific optogenetic stimulation in rhesus macaques. *Cell* 166(6):1564–71.e6

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. 2013. A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* 16(7):966–73

Stolz DS, Müller-Pinzler L, Krach S, Paulus FM. 2020. Internal control beliefs shape positive affect and associated neural dynamics during outcome valuation. *Nat. Commun.* 11(1):1230

Stopper CM, Tse MT, Montes DR, Wiedman CR, Floresco SB. 2014. Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron* 84(1):177–89

Syed ECJ, Grima LL, Magill PJ, Bogacz R, Brown P, Walton ME. 2015. Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat. Neurosci.* 19(1):34–36

Tang S, Cui L, Pan J, Xu N-L. 2024. Dynamic ensemble balance in direct- and indirect-pathway striatal projection neurons underlying decision-related action selection. *Cell Rep.* 43(9):114726

Taverna S, Ilijic E, Surmeier DJ. 2008. Recurrent collateral connections of striatal medium spiny neurons are disrupted in models of Parkinson's disease. *J. Neurosci.* 28(21):5504–12

Thomas AW, Molter F, Krajbich I. 2021. Uncovering the computational mechanisms underlying many-alternative choice. *eLife* 10:e57012

Tsutsui-Kimura I, Matsumoto H, Akiti K, Yamada MM, Uchida N, Watabe-Uchida M. 2020. Distinct temporal difference error signals in dopamine axons in three regions of the striatum in a decision-making task. *eLife* 9:e62390

Turner RS, Desmurget M. 2010. Basal ganglia contributions to motor control: a vigorous tutor. *Curr. Opin. Neurobiol.* 20(6):704–16

Urakubo H, Yagishita S, Kasai H, Kubota Y, Ishii S. 2021. The critical balance between dopamine D2 receptor and RGS for the sensitive detection of a transient decay in dopamine signal. *PLOS Comput. Biol.* 17(9):e1009364

van der Schaaf ME, van Schouwenburg MR, Geurts DE, Schellekens AF, Buitelaar JK, et al. 2014. Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning. *Cereb. Cortex* 24(3):633–42

Voon V, Pessiglione M, Brezing C, Gallea C, Fernandez HH, et al. 2010. Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron* 65(1):135–42

Wang Y, Lak A, Manohar SG, Bogacz R. 2024. Dopamine encoding of novelty facilitates efficient uncertainty-driven exploration. *PLOS Comput. Biol.* 20(4):e1011516

Watabe-Uchida M, Eshel N, Uchida N. 2017. Neural circuitry of reward prediction error. *Annu. Rev. Neurosci.* 40:373–94

Wenzel JM, Oleson EB, Gove WN, Cole AB, Gyawali U, et al. 2018. Phasic dopamine signals in the nucleus accumbens that cause active avoidance require endocannabinoid mobilization in the midbrain. *Curr. Biol.* 28(9):1392–404.e5

Westbrook A, Frank MJ, Cools R. 2021. A mosaic of cost-benefit control over cortico-striatal circuitry. *Trends Cogn. Sci.* 25(8):710–21

Westbrook A, van den Bosch R, Määttä JI, Hofmans L, Papadopetraki D, et al. 2020. Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science* 367(6484):1362–66

Wiecki TV, Frank MJ. 2013. A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychol. Rev.* 120(2):329–55

Wiecki TV, Riedinger K, von Ameln-Mayerhofer A, Schmidt WJ, Frank MJ. 2009. A neurocomputational account of catalepsy sensitization induced by D2 receptor blockade in rats: context dependency, extinction, and renewal. *Psychopharmacology* 204(2):265–77

Wietzikoski EC, Boschen SL, Miyoshi E, Bortolanza M, et al. 2011. Roles of D1-like dopamine receptors in the nucleus accumbens and dorsolateral striatum in conditioned avoidance responses. *Psychopharmacology* 219(1):159–69

Yagishita S, Hayashi-Takagi A, Ellis-Davies GC, Urakubo H, Ishii S, Kasai H. 2014. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345(6204):1616–20

Yartsev MM, Hanks TD, Yoon AM, Brody CD. 2018. Causal contribution and dynamical encoding in the striatum during evidence accumulation. *eLife* 7:e34929

Yttri EA, Dudman JT. 2016. Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* 533:402–6

Zachry JE, Kutlu MG, Yoon HJ, Leonard MZ, Chevée M, et al. 2024. D1 and D2 medium spiny neurons in the nucleus accumbens core have distinct and valence-independent roles in learning. *Neuron* 112(5):835–49.e7

Zalocusky KA, Ramakrishnan C, Lerner TN, Davidson TJ, Knutson B, Deisseroth K. 2016. Nucleus accumbens D2R cells signal prior outcomes and control risky decision-making. *Nature* 531:642–46