In: Jaeger D., Jung R. (Ed.) *Encyclopedia of Computational Neuroscience*. Springer, 2013.

Computation with Dopaminergic Modulation

Samuel J. Gershman
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Definition

Dopamine (DA) is a neuromodulator released by midbrain neurons with widespread projections throughout the brain. Dopaminergic modulation has diverse effects on cellular, motor and cognitive functions, including reinforcement learning, working memory and attention. Dysregulation of dopamine also plays a central role in the breakdown of these functions in disorders such as Parkinson's disease and schizophrenia.

Detailed Description

Dopamine basics

DA is a neuromodulator released by neurons in two nuclei of the midbrain: the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc). The axons of these neurons project to a large number of cortical and subcortical areas. Prefrontal cortex, motor cortex, and the striatum are among the most densely innervated areas. DA release tends to be fairly homogeneous across DA neurons, by virtue of electrical coupling between the axons of adjacent neurons, which induces highly synchronous firing. As pointed out by Paul Glimcher, these properties suggest that DA neurons "cannot say much to the rest of the brain but what they say must be widely heard" (Glimcher, 2011).

DA affects its targets primarily by two classes of receptor—D1 and D2—and the distinct properties of these receptors have functional consequences which are discussed below. D2 receptors tend to be activated earlier, and at lower concentrations, than D1 receptors (Lapish et al., 2007; Schultz, 2007). As a consequence, D2 receptors are more sensitive to phasic (transient) DA release than D1 receptors, with the opposite pattern for tonic (background) DA release (Grace, 1991). D1 and D2 receptors also have different postsynaptic effects: D1 receptor activation increases both NMDA (excitatory) and GABA (inhibitory) currents, whereas D2 receptor activation decreases them (Trantham-Davidson et al., 2004).

Reinforcement learning models

One of the most influential hypotheses about DA is that it plays an integral role in a system that learns to predict future reward. This hypothesis can be formalized in terms of the theory of reinforcement learning (RL) first developed in computer science (Sutton & Barto, 1998). At each time *t* the agent occupies a state s_t (e.g., the agent's location or the surrounding stimuli) and receives a reward r_t. The goal of an RL system is to estimate the *expected discounted future return*, or *value*, of visiting a sequence of states starting in state s_t:

$$V(s_t) = E[\sum_{k=0} \gamma^k r_{t+k}],$$

where γ is a *discount factor* that account for the fact that distal rewards are less valuable than proximal rewards. The expectation *E* represents an average of the quantity inside the brackets over stochastic sequences of states.

If we make certain assumptions about the environment (specifically, that it is a *Markov decision process*; see Sutton & Barto, 1998), then the value can be estimated by a simple learning rule known as *temporal difference learning*:

$$\hat{V}(s_t) \leftarrow \hat{V}(s_t) + \alpha\delta_t,$$

where α is a *learning rate* and δ_t is a *prediction error* defined as:

$$\delta_t = r_t + \gamma\hat{V}(s_{t+1}) - \hat{V}(s_t).$$

Intuitively, if the prediction error is greater than zero, then more reward was received than was predicted, and hence $\hat{V}(s_t)$ is increased. If the prediction error is less than zero, then $\hat{V}(s_t)$ is decreased. Another implication of this learning rule is that prediction errors will gradually propagate backwards in time as values at earlier time points come to reflect reward received at later time points.

Schultz, Dayan and Montague (1997) observed that the phasic firing of DA neurons in an appetitive conditioning experiment was conspicuously consistent with a hypothetical prediction error. In these experiments, a conditioned stimulus (CS; e.g., a tone) is repeatedly followed by an unconditioned stimulus (US; e.g., juice reward). Schultz et al. noticed that DA firing increased in response to an unexpected reward (e.g., in early training trials). After multiple training trials, if the reward is delivered following the CS, then DA firing is at baseline following reward delivery but increases following the CS onset. In this case, the reward is expected and hence the prediction error at the time of reward is zero, whereas the prediction error at the time of the CS is greater than zero due to the temporal propagation described above. Finally, if the reward is omitted following the CS, DA firing pauses (decreases below baseline), consistent with the occurrence of a negative prediction error.

In addition to its role in predicting reward, DA plays an important role in selecting actions to increase reward. For example, electrophysiological studies have shown that DA responses to

reward are sensitive to actions (Morris et al, 2006; Roesch et al., 2007), and DA receptor antagonism in the striatum suppresses the initiation of reward-seeking actions (Berridge & Robinson, 1998). According to RL theory, there are a number of ways to use prediction errors to estimate the value of actions, some of which have been integrated into models of DA (e.g., Frank, 2005; Houk et al., 1995; Montague et al., 1996; Suri & Schultz, 1998).

Since the seminal work of Schultz et al. (1997), the prediction error theory of DA has been corroborated by converging evidence from humans and other animals (see Glimcher, 2011 for a review). At the same time, it is clear that such a simple theory is unable to account for many of the complexities of DA function. These lacunae have stimulated a number of significant extensions of the theory, some of which are described below. Interestingly, these extensions mirror the technical development of RL theory in computer science. Biological and artificial agents face similar challenges in optimizing reward—and may have even arrived at similar solutions.

One development of the prediction error theory concerns the phasic DA firing in response to novel stimuli (e.g., Horvitz et al., 1997). The original prediction error theory fails to anticipate this finding, since novelty does not by itself predict reward. However, as pointed out by Kakade and Dayan (2002), novelty responses can be rationalized by considering the problem of optimizing reward. All agents face a dilemma between *exploiting* the currently best option and *exploring* other options that might be potentially better (Cohen et al., 2007). One way to encourage exploration is by initializing the values of all states to be greater than zero. As shown by Kakade and Dayan, this has the effect of generating positive prediction errors when these states are visited for the first time, consistent with the DA novelty response.

A second development has been to incorporate timing variability and partial observability into the TD model (Daw et al., 2006). The original prediction error theory (Montague et al., 1996; Schultz et al., 1997) used a delay line to represent different temporal epochs as distinct states. Daw et al. (2006) replaced the delay line with two assumptions: (1) States are occupied for some random amount of time before a transition occurs (timing variability); and (2) states are not directly observed, but must be inferred from noisy sensory data (partial observability). Daw et al. showed that these extensions could account for a number of timing properties of phasic DA firing. For example, if a reward is delivered early or late, a phasic DA burst is observed (Hollerman & Schultz, 1998), contrary to the delay line model, which predicts a negative prediction error (i.e., a pause in firing). The Daw et al. model correctly predicts a positive prediction error, because it infers that a transition to the “reward state” has occurred early or late.

A third development of the theory concerns the role of tonic DA. It has long been recognized that tonic levels of DA appear to be regulated independently of phasic levels (Grace, 1991). In particular, increased tonic DA has been associated with the invigoration of behavior (Schultz, 2007), and the depletion of tonic DA results in a loss of vigor (Salamone et al., 2001; Sokolowski & Salamone, 1998). To account for these observations, Niv et al. (2007) proposed that tonic DA signals an estimate of an environment's average reward. When the average reward is high, it is rational for an agent to respond more vigorously (since more reward can be collected), whereas when the average reward is low, the energetic costs may outweigh the benefits of responding. This model predicts not only a lower rate of responding with depleted tonic DA, but also a longer latency to respond, consistent with experimental evidence (Denk et al., 2005).

A fourth development has been to integrate the prediction error theory into more biologically detailed models of RL in the basal ganglia. For example, Frank (2005) proposed a model in which DA differentially affects the "direct" and "indirect" pathways of the basal ganglia. The direct pathway Striatal neurons in the direct pathway primarily express D1 receptors, whereas striatal neurons in the indirect pathway primarily express D2 receptors. The net effect of DA in the striatum is to increase activity in the direct pathway and suppress activity in the indirect pathway, resulting in disinhibition of action selection in the frontal cortex via modulation of basal ganglia output pathways. One consequence of this differential effect is that neurons in the direct pathway undergo long-term potentiation following a DA burst, whereas neurons in the indirect pathway undergo long-term depression. Another contribution of this model is a division of error-driven learning into positive and negative components: D1-expressing neurons learn more from positive prediction errors, whereas D2-expressing neurons learn more from negative prediction errors. Thus, the model of Frank (2005) provides a neurally plausible mechanism by which DA can influence action selection in corticostriatal circuits.

Gain modulation and working memory

The sensitivity of a neuron to stimulation can be enhanced by the application of DA, without changing the spontaneous firing rate (Clarke et al., 1987; Foote & Morrison, 1987). In other words, DA increases the signal-to-noise ratio of a neuron. Servan-Schreiber, Printz and Cohen (1990) introduced a connectionist implementation of this idea by positing that DA increases the gain β of a sigmoidal activation function:

$$a = \frac{1}{1+e^{-\beta x - b}},$$

where *x* is the synaptic input of the neuron, *b* is the baseline membrane potential, and *a* is the firing rate. This basic schema for dopaminergic gain modulation has been reproduced in much greater biological detail by a number of authors (e.g., Durstewitz et al., 1999; Moyer, Wolf & Finkel, 2007).

An important function of gain modulation is to support active maintenance in working memory: If a group neurons are recurrently connected to form an attractor network, increasing the gain will increase the stability of high activity states (i.e., actively maintained memories) while suppressing low activity states corresponding to spontaneous background activity. Consistent with this idea, DA appears to play an important role in supporting active maintenance of working memory representations in the prefrontal cortex (Cohen, Braver & Brown, 2002).

DA has been hypothesized to play a complementary role in updating working memory (Braver & Cohen, 1999; Durstewitz et al., 1999; O'Reilly & Frank, 2006). According to these models, a phasic burst of DA acts as a "gating" signal, transiently amplifying afferent inputs while suppressing local inhibitory signals. The effect of this gating signal is the encoding of afferent input into the current attractor state. O'Reilly and Frank (2006) have proposed a biologically detailed refinement of this idea, whereby "stripes" in the prefrontal cortex (groups of neurons connected to distinct areas of the basal ganglia. Levitt et al. (1993) encode representations in working memory that are updated by phasic DA bursts. This model uses an RL mechanism to learn when to gate, providing a link to the models of corticostriatal plasticity described above.

How can the active maintenance and updating roles of DA be reconciled? One suggestion is that these roles are mediated by different receptors. Specifically, Cohen et al. (2002) have suggested that phasic DA activity primarily affects updating via D2 receptors, whereas tonic DA activity primarily affects active maintenance via D1 receptors. These two modes interact, with tonic DA inhibiting phasic DA by regulating presynaptic autoreceptors (Grace, 1991). This hypothesis is consistent with the observation that prolonged pharmacological elevation of tonic DA results in perseverative and stereotyped behavior (Arnsten, 1997; Sawaguchi & Goldman-Rakic, 1991), which could arise from a failure to update due to reduced phasic DA.

Computational modeling of dopamine-associated disorders

The models reviewed above furnish a set of hypotheses about the etiology of psychiatric disorders arising from DA dysfunction. This section focuses on the two most well-studied examples: schizophrenia and Parkinson's disease.

A number of authors have proposed that many cognitive deficits in schizophrenia can be explained by a breakdown in the regulation of gating (Braver & Cohen, 1999; Durstewitz & Seamans, 2008; Frank, 2008; Waltz et al., 2007). The gain modulation view lead Braver and Cohen (1999) to propose that reduced prefrontal DA in schizophrenics results in a lower signal-to-noise ratio; combined with noisier DA signaling, this produces gating of irrelevant

information into working memory, and can cause psychotic episodes in schizophrenia. Another view, proposed by Frank and colleagues (Frank, 2008; Waltz et al., 2007), is that striatal D1 receptor function is compromised in schizophrenia, resulting in impaired learning from positive prediction errors but spared learning from negative prediction errors.

The model of Frank and colleagues also sheds light on the cognitive deficits in Parkinson's disease (Frank, 2005; Moustafa et al., 2008b; Wiecki & Frank, 2010). Unmedicated patients with Parkinson's disease appear to be better at learning from negative outcomes compared to positive outcomes, and this pattern is reversed by DA medication (Cools et al., 2006; Frank et al., 2004; Moustafa et al., 2008a). Frank et al. explained these findings in terms of the effects of DA levels on D1 and D2 receptors in the striatum: Depleted DA in unmedicated patients results in D2 dominance (as in schizophrenia), whereas elevated DA in medicated patients results in D1 dominance. This account may explain the nature of cognitive deficits following DA medication, such as failures to avoid non-rewarding or punishing stimuli (Cools et al., 2006; Frank et al., 2004; Moustafa et al., 2008a) and susceptibility to pathological gambling and addiction (Dagher & Robbins, 2009).

In addition to these cognitive deficits, Parkinson's disease is associated with bradykinesia, an overall slowing of movements, and this deficit is relieved by DA medication. However, Parkinson's patients can learn to make movements at a task-specified velocity and with the same level of accuracy as controls; the key difference is that Parkinson's patients take longer to reach the performance criterion (Mazzoni et al., 2007). The tonic DA model of Niv et al. (2007) offers a perspective on these findings: since tonic DA is reduced in Parkinson's disease, the average reward will be perceived as lower, thus reducing the opportunity cost of not responding. Thus, bradykinesia may actually reflect an optimal decision by the motor system based on incorrect information about average reward provided by tonic DA (Niv & Rivlin-Etzion, 2007).

References

Arnsten AFT (1998) Catecholamine modulation of prefrontal cortical cognitive function. Trends Cogn. Sci. 2:436-447

Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res. Rev. 28:309–369

Braver TS, Cohen JD (1999) Dopamine, cognitive control, and schizophrenia: the gating model. Prog. Brain Res. 121:327-349

Clarke CR, Geffen GM, Geffen LB (1987) Catecholamines and attention I: animal and clinical studies. Neuroscience and Biobehavioral Reviews 11:341-352

Cohen JD, Braver TS, Brown JW (2002) Computational perspectives on dopamine function in prefrontal cortex. Current Opinion in Neurobiology 12:223-229

Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 362:933–942

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

Dagher A, Robbins T (2009) Personality addiction dopamine: Insights from Parkinson's disease. Neuron 61:502–510

Daw ND, Courville AC, Touretzky DS (2006) Representation and timing in theories of the dopamine system. Neural Computation 18:1637–1677

Denk F, Walton ME, Jennings KA, Sharp T, Rushworth MF, Bannerman DM (2005) Differential involvement of serotonin and dopamine systems in cost–benefit decisions about delay or effort. Psychopharmacology (Berl) 179:587–596

Durstewitz D, Kelc M, Güntürkün O (1999) A neurocomputational theory of the dopaminergic modulation of working memory functions. Journal of Neuroscience 19:2807-2822

Durstewitz D, Seamans JK (2008) The dual-state theory of prefrontal cortex dopamine function with relevance to COMT genotypes and schizophrenia. Biological Psychiatry 64:739-749

Foote L, Morrison JH (1987) Extrathalamic modulation of cortical function. Annual Review of Neuroscience 10:67-95

Frank MJ, Seeberger L, O'Reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. Science 306:1940-1943

Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. Journal of Cognitive Neuroscience 17:51–72

Frank MJ (2008) Schizophrenia: A computational reinforcement learning perspective. Schizophrenia Bulletin 34:1008-1011

Glimcher PW (2011) Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proceedings of the National Academy of Sciences. 108:15647-15654

Grace AA (1991) Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis of the etiology of schizophrenia. Neuroscience. 41:1-24

Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience 1:304-309

Horvitz JC, Stewart T, Jacobs B (1997). Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. Brain Research. 759:251–258

Houk JC, Adams JL, Barto, AG (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In JC Houk, JL Davis, DG Beiser (Eds.), Models of information processing in the basal ganglia (pp. 249–270). Cambridge, MA: MIT Press

Kakade S, Dayan P (2002) Dopamine: generalization and bonuses. Neural Networks. 15:549-559

Lapish CC, Kroener S, Durstewitz D, Lavin A , Seamans JK (2007) The ability of the mesocortical dopamine system to operate in distinct temporal modes. Psychopharmacology. 191:609–625

Levitt JB, Lewis DA, Yoshioka T, Lund JS (1993). Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). Journal of Comparative Neurology 338:360–376

Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 27:7105–7116

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. Journal of Neuroscience 16(5):1936-47

Moustafa AA, Cohen MX, Sherman SJ, Frank, MJ (2008a) A role for dopamine in temporal decision making and reward maximization in parkinsonism. Journal of Neuroscience, 28:12294–12304

Moustafa AA, Sherman SJ, Frank MJ (2008b) A dopaminergic basis for working memory, learning, and attentional shifting in parkinson's disease. Neuropsychologia 46:3144–3156

Moyer JT, Wolf JA, Finkel LH (2007) Effects of dopaminergic modulation on the integrative properties of the ventral striatal medium spiny neuron. Journal of Neurophysiology 98:3731-48

Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl) 191:507–520

Niv Y, Rivlin-Etzion M (2007) Parkinson's disease: Fighting the will? Journal of Neuroscience 27:11777–11779

O'Reilly RC, Frank MJ (2006) Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. Neural Computation 18:283–328

Salamone JD, Wisniecki A, Carlson BB, Correa M (2001) Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement. Neuroscience 5:863–870

Sawaguchi T, Goldman-Rakic PS (1991) D1 dopamine receptors in prefrontal cortex: involvement in working memory. Science 251:947-950

Servan-Schreiber D, Printz H, Cohen JD (1990) A network model of catecholamide effects: gain, signal-to-noise ratio, and behavior. Science 249:892-895

Schultz W, Dayan P and Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599

Schultz W (2007) Multiple dopamine functions at different time courses. Annual Review of Neuroscience 30:259–88

Sokolowski JD, Salamone JD (1998) The role of accumbens dopamine in lever pressing and response allocation: effects of 6-OHDA injected into core and dorsomedial shell. Pharmacol

Biochem Behav 59:557–566

Suri RE, Schultz W (1998) Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Experimental Brain Research 121:350–354

Sutton RS, Barto AG (1998) Reinforcement Learning: an Introduction. Cambridge, MA: MIT Press

Trantham-Davidson H, Neely LC, Lavin A, Seamans JK (2004) Mechanisms underlying differential D1 versus D2 dopamine receptor regulation of inhibition in prefrontal cortex. Journal of Neuroscience 24:10652-10659

Waltz JA, Frank MJ, Robinson BM, Gold JM (2007) Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. Biol. Psychiatry. 62:756–764

Wiecki TV, Frank MJ (2010) Neurocomputational models of motor and cognitive deficits in Parkinson's disease. Progress in Brain Research 183:275-297