Dopamine modulation in basal ganglia-cortical network implements saliency-based gating of working memory

Aaron Gruber[1,2], Peter Dayan[3], Boris Gutkin[3], and Sara Solla[2,4]

Biomedical Engineering[1], Physiology[2], and Physics and Astronomy[4] Northwestern University, Chicago, IL
{ *a-gruber1, solla* }*@northwestern.edu*

Gatsby Computational Neuroscience Unit[3] University College London, London, UK
{ *dayan, boris* }*@gatsby.ucl.ac.uk*

Abstract

The sustained neural activity in prefrontal cortex (PFC) underlying working memory (WM) is under under control of dopamine (DA) both through direct modulatory effect on pyramidal neurons and also through the control of basal ganglia (BG) input to PFC. We construct a network model of BG-PFC interaction to investigate the combined influence of DA on WM formation through both pathways. The cortical component is a line attractor which maintains memory of input through a persistent 'bump' of activity localized to a subgroup of PFC units. DA increases the contrast of these units leading to increased robustness of the encoded quantity, similar to results of other models of cortical DA effects on WM. The activation of medium spiny neurons (MSN) of the BG provides an additional input to the cortex, which increases the robustness of the memory to noise and also gates access to working memory by triggering input driven switches of the bump location. Our previously reported biophysically-grounded model of MSNs shows that these neurons become bistable in elevated dopamine, which accounts for both the suppressive and enhancing effects of DA. DA-induced bistability provides a novel mechanism for the network to selectively gate input to WM based on salience: inputs that elicit DA response are strongly encoded in WM through enhanced MSN activation, but subsequent inputs are blocked from switching the WM while DA is elevated through a suppression of MSN activity to these inputs. The combined actions of DA in the BG and PFC lead to WM that is selective for salient input and that has increased robustness to distraction.

1 Introduction

Experimental evidence indicates that the maintenance of information in WM is mediated by persistent neural activity in the PFC. Modulation of this activity ei-

ther through neuromodulation of cortical response properties or alteration of the input to PFC provides mechanisms for controlling WM. Many computational models of DA effects on WM focus on direct neuromodualtion in the cortex, which has been shown experimentally to influence WM. Unfortunately, the multifaceted effects of DA modulation on the electrophysiology of neurons are not fully understood, complicating the construction of computational models. A recent cellular model including many known effects of DA on ionic conductances indicates that DA modulation of pyramidal neurons causes the pattern of network activity at a fixed point attractor to become more robust to noise and input -driven switching of attractor states [7], consistent with the properties of spiking models investigating effects of DA on NMDA and AMPA-mediated transmission and is effects on fixed point attractors [2]. Interestingly, many different styles of models, form high level models designed to investigate human psycobehavioral data [**?**], to biophysically motivated cellular models [**?**] and intermediate models [**?**], broadly share a common phenomenology: DA reduces the efficacy of input for switching WM representation, and in many models, also enhances interaction between active units. Our model indicates that both input from BG and its modulation by DA provide additional functionality to WM.

In addition to modulatory inputs, both excitatory and inhibitory input to PFC also control WM formation. A prominent input to PFC comes from the BG. Although the BG has played a prominent role in models of motor control, this pathway has received very limited attention in models of WM. One group [3] has constructed network models for cognitive tasks that require working memory has ascribed a gating function to the BG in the absence of DA in which the BG disinhibits selective PFC units allowing activation of those units, an idea that has been long-standing in the motor control community. Our goal in this work is to investigate how a biologically-grounded model of BG gates information to PFC, and how the complex effect of DA on MSNs will affect this gating function. We choose to model the WM demand for the memory guided saccade task that is commonly used to study neural activity involved in WM. In this task, subjects must fixate a centrally located spot while a briefly illuminated visual target located in a circular position around a fixation point is presented. Following a delay period of a few seconds, subjects must saccade to the remembered target location. We use a 'bump' network (line attractor) similar to those used by others [1, 4] to model the working memory of the target angle on the circle. An advantage of this task and model formulation is that the wealth of single unit data from animals performing this task can be used to construct the model and to test its validity. One intriguing finding is that the activity of striatal MSNs show a profound reward dependence [6]. We have shown in a previous model that this response modulation could be mediated by DA [8], providing a route for controlling WM formation in addition to direct modulatory effects in the PFC. We show here that the additional gating imparted by the BG and its modulation by DA work in a complementary manor with DA modulation in the cortex to gate in and enhance WM of salient information.

2 Model description

3 Results

Dopamine effects in the cortex: increased memory robustness

We first investigate the cortical network in isolation from the input and BG components. The connectivity of the cortical units is set so as to allow two stable conditions for excitatory cortical network activity in the absence of external input

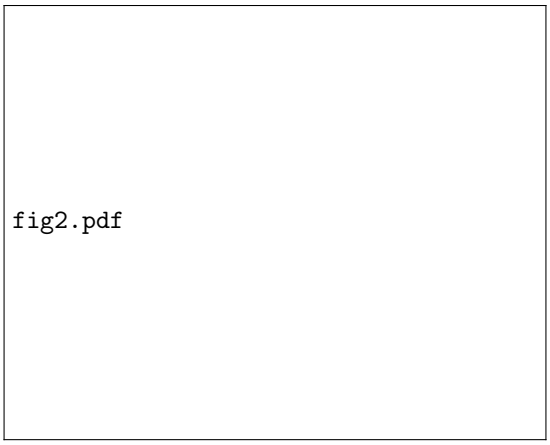

Figure 2: Activity profile of bump state in low DA (open circles) and high DA (solid circles) (A). The activity of cortical units is shown in pseudo-color versus time with white as most active (B, inset) showing the influence of cortical injection at $\theta^{\dagger}$ in moving the bump from θ_0 to θ_1. The plot in B shows the sensitivity of the bump over injection locations from 0 to π in low DA (dotted line) and high DA (solid line).

and noise: either all units have very low activity, or a subset of units participate in a 'bump' of elevated activity shown by the filled dots in Fig 2A. This configuration is capable of encoding a continuous angle by assigning each unit a direction label and translating the bump on the ring of cortical units; the encoded angle is the corresponding labeled direction at the peak of the bump, and can be read out by computing the population vector. The action of DA in the cortex is modeled as increasing the slope of the logistic function of the cortical units, leading to a more narrow activity bump with a higher peak as shown by the open dots in Fig 2A. This leads to an increased stability of the bump location, and therefore encoded angle, as described next.

External excitatory input is capable of moving the bump. A quantitative measure of bump robustness against movement is obtained by providing uniform input to five units for a fixed duration and measuring the resulting bump displacement in the absence of noise. This procedure is illustrated in figure 2B. The activity of the cortical units is displayed in pseudo-color versus time to visualize network dynamics. Current injection centered at $\theta^{\dagger}$ causes a drift of the bum initially centered at (or encoding?) θ_0 towards the injection location. The bump ceases to move once the injection is terminated, and its location θ_1 is read out 40 ms later. Bump displacement is dependent on the location of the injection relative to the bump location as shown in Fig 2C. The injection is capable of moving the bump completely to the injection location when the injection occurs close to θ_0, however, the injection becomes less effective as the distance increases. The incomplete bump movement illustrated in Fig 2B represents the data point in Fig 2C marked by an aserisk. The intensity and duration of injection affect the sensitivity profile, but the trends remain qualitatively similar. The increases in activation function slope in high DA decreases the sensitivity of the bump to current injection (Fig 2C, dotted line), hence increasing the robustness of the encoded quantity. DA also reduces the diffusion of the bump caused by noise as shown in Fig 3. The trajectories (inset) and variance of the population vector readout for 60 independent realizations are displayed in conditions of low (dashed line) and high (solid line) cortical DA. The increased robustness of the memory encoded by the bump in high DA is consistent with reported results obtained with other line and fixed point attractor models of DA effects on WM.

BG effects on cortex: increased memory robustness and gating actions

We now investigate the effects of BG input on cortical bump stability in the absence of dopamine modulation. Tonic input from a single MSN to the cortical units is

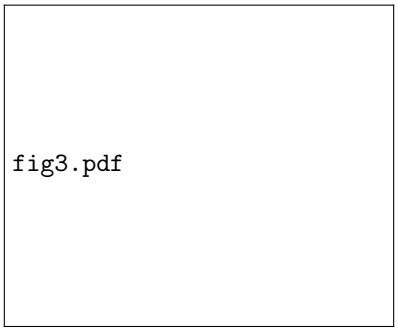

Figure 3: Effects of DA on bump diffusion. Sample traces of the population vector in low DA (inset) and its variance in low DA (dotted line) and high DA (solid line)

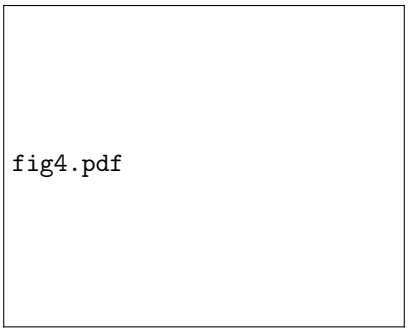

Figure 4: Noise driven bump diffusion that occurs in low DA (A, dotted line) is reduced by input from a singe BG unit (A, solid line). The robustness to switching is also increased by this input (B)

able to anchor the location of the bump, leading to an increased robustness of the bump to noise induced diffusion as shown in Fig 4A, and to current injection as shown in Fig 4B. The tonic input to cortex has the effect of breaking the symmetry of the line attractor in the cortex and causing a single fixed point occurring at the location of maximal input to the cortex, which corresponds to the preferred direction of the active MSN (but the cortical net can still go to 0 activity, so its still bistable, but with the active state being a fixed point rather than line attractor). The transition to a single fixed point reduces the maximal deviation of the bump due to current injection, and also has the property of pulling the bump back to the fixed point following a perturbation. Therefore θ_1 is computed with short latency (40 ms) after injection offset so that the deviation is reflected in the plot in Fig 4B before reconvergence to the fixed point.

The output of the MSNs provides control over the encoded memory not only by enhancing robustness, but also by providing input that can assist the direct visual input to the cortex in triggering a switch in the attractor location when a visual stimulus is presented. The network activity resulting from the presentation of two different target locations is shown in Fig 5. The top plots show the angle encoded by a gaussian bump of activity in the input layer as a grey bar, superimposed with a thin black line indicating the population vector of the cortical layer. The middle and bottom plots show the activity of BG and PFC layers respectively. The initial input activates one MSN strongly and later two neighboring units, and causes activation of a group of PFC units. When the input is extinguished, the MSNs become inactive and the cortical layer relaxes to its characteristic bump state. In Fig 5A, the presentation of a second input stimulus does not activate MSN units and has only a minimal effect on the bump state. In Fig 5B, the second input activates MSNs and the resulting additional input to the cortex causes a switch of the bump state so that it encodes the new stimulus location. The MSNs are effectively gating the input to the cortex and can control access to working memory.

DA effects in BG: saliency-based gating

We now explore the effects of DA on the properties of MSNs and the resulting effects on the gating function of BG. Many lines of evidence indicate that DA modulation has an impact on neuronal activity of MSNs. Data of particular relevance to our

Figure 5: Cortical input to PFC in not sufficient to switch attractor states (A). Input driven MSN activation triggers switching (B), gating information into WM

model comes from single unit recordings of MSNs in monkeys performing a memory guided saccade task. The responses of many MSNs to visual targets were of higher frequency and longer duration when the targets indicated reward rather than an audible tone for correct task performance. A minority of MSNs were suppressed in the reward condition. Visual stimuli indicating reward are known to trigger DA response. Our model of MSNs accounts for both types of response modulation through DA-induced bistability. We identified two consequences of DA-induced bistability that can have a significant impact on the response and potentially the function of MSNs: firstly, the threshold for reaching the active 'up' state is increased by DA, but secondly, units that reach the up state have higher frequency activity that can be extended in duration due to hysteresis.

The dual enhancing/suppressing nature of DA modulation on MSNs has significant consequences on the network's response to inputs, as shown in Fig 6. The top 4 panels show the input angle superimposed with the encoded angle, DA signal γ, BG activity, and PFC activity respectively. The sequence of inputs is θ_A, θ_B, θ_A^*, θ_B, where θ_A^* is a conditioned stimulus that triggers DA release. The MSNs activate following the first two inputs and allow the input to be encoded in working memory. The presentation of θ_A^* begins to activate the same set of MSNs as θ_A but the units become bistable allowing only one unit to become active, consequently switching the bump state to encode θ_A^*. Now however, the presentation of θ_B does not elicit MSN activation because the total excitatory input is below the threshold for reaching the up state while MSNs are bistable, and the bump state continues to encode θ_A^*. DA modulation in the BG has imparted salience selectivity to the gating function of

Figure 6: DA locks the activation of MSNs, preventing a switch to stimuli following DA release and preserving the conditioned stimulus in WM.

the BG by inducing bistability. In addition, the robustness of the bump during the delay period is enhanced by the combined effect of DA through both increasing the slope of cortical unit activation function and by sustaining MSN input during the delay (Fig 6, bottom panel; dashed line compared to solid line).

4 Conclusions

Our model represents one of the first efforts to link DA modulation in both cortex and BG and the resulting influence of BG input to cortex on WM formation. Our model indicates that both input from BG and its modulation by DA complements the cortical effects of DA to provide more sophisticated control over WM and addresses some of the fundamental drawbacks inherent to line attractor models.

Diffusion: Line attractor models are appealing for many types of WM demands such at the saccade task since they can encode a continuum of values. A consequence of this ("attractive") property is the susceptibility of the encoded quantity to diffusion-like distortion due to noise. Fixed point attractors do not have this property because the attractor dynamics oppose the noise induced perturbations and the encoded value does not accumulate noise as the line attractor does. The BG input effectively creates fixed points in the cortical network through MSN activation, essentially eliminating the diffusion (Fig 4A). We also find that DA can decreases the diffusion through modulation in the PFC by increasing the slope of

cortical activation functions (FIg 3). Camperi and Wang show that bistable cortical units reduce the diffusion. However, significantly large noise will exceed the bistable region and lead to diffusion. Significantly large noise in our model could also kick the network out of the basin of attraction and corrupt the memory, but the attractor point will remain as long as MSNs are active.

Gating: Many computational models ascribe a gating function to the effects of DA on cortical neurons. The biologically-grounded models of DA modulation investigate the stability of fixed point attractors [**?**], and indicate that attractor points become more robust to switching. These models, as well as other more abstract models of DA function in the cortex, seem to broadly share a common phenomenology: units that are inactive are less excitable by input, while units that are active can become more active. Although there are some subtle and important differences, these effects are generally consistent with the more abstract notion that the gain (or contrast) of cortical units is increased by DA [5]. This gain increase is also apparent at intermediate levels of DA modulation of our MSN model, before the unit becomes bistable. The nature of DA modulation seems consistent in BG and PFC; it is the intensity of modulation that may be more pronounced in BG leading to bistability. Our model indicates that the more abstract effects of DA implemented in the cortical model (contrast enhancement) increase the robustness of the WM, consistent with other interpretations of DA function. Many models ascribe a gating function to this DA-mediated stabilization of cortical activity that is critical for WM. However, this places a heavy workload on phasic DA release: it must respond constantly on short time intervals to control information flow into cortex. We have shown here that the BG can impart gating of information into cortex in low DA by controlling the switching of the attractor state in response to inputs. BG-mediated gating has the advantage of being likely more temporally precise and spatially specific than a modulator system is likely to be - both advantageous properties for WM control (too speculative?).

The idea of BG mediated facilitation of cortical activity has been prominent in motor control communities, but has only recently been incorporated in cognitive models. Frank et al have formulated a network model of punctate neurons in which the BG triggers a switch in the PFC units to which they project, to that the PFC units become bistable and maintain any activity induced by input from other cortical areas. This model is able to perform a difficult task requiring WM, but does not reflect the notion that DA modulation can signal saliency of sensory information. This notion is emphasized by experimenters examining the reward dependence of MSN single unit activity [6] and animal behavior (other Hikosaka ref or consolidate with Kawagoe98?). These authors emphasize that the BG is at the confluence of converging sensory information from cortex and motivationally-related DA signal. Our model is an attempt to model the mechanisms underlying the effects of DA modulation on BG activity and the influence of the modulated output of the BG on WM. DA-induced bistability of MSNs imparts a salience selectivity of the BG gating function in part by nonlinearly increasing the threshold for MSNs to reach the active up state. It is the Da-Controlled threshold in conjunction with the amount of excitatory input to MSNs that control activity of MSNs. Since DA influences the modification of striatal synapses, it controls the input to MSN and thus the conditions under which this threshold is exceeded. This form of learning, in conjunction with DA-induced bistability of MSN, enhances the BG's competence at saliency-based gating.

Our model indicates a biologically plausible mechanism for the role of BG and DA in gating access to PFC and influencing WM. This control pathway enhances the effects of DA in the cortex to provide a more sophisticated control over WM to both

gate information and increase its stability.

References

[1] Camperi M, Wang XJ (1998). *J. Comp. Neurosci.* 5:383-405.

[2] Brunel N, Wang XJ (2001). *J. Comp. Neurosci.* 11:63-85.

[3] Frank M, Loughry B, O'Reilly (2001). *Cog., Affective, & Behav. Neurosci.* 1(2):137-160.

[4] Compte A, Brunel N, Goldman-Rakic P, Wang XJ (2000). *Cereb. Cortex* 10:910-923.

[5] Servan-Schreiber D, Printz H, Cohen J (1990). *Sci.* 249:892-895.

[6] Kawagoe R, Takikawa Y, Hikosaka O (1998). *Nat, Neurosci.* 1:411-416.

[7] Durstewitz D, Seamans J, Sejnowski T (2000). *J, Neurophysiol.* 83:1733-1750.

[8] Gruber A, Solla S, Houk JC (2003). *NIPS*

[9] Amit D, Brunel N (1996)