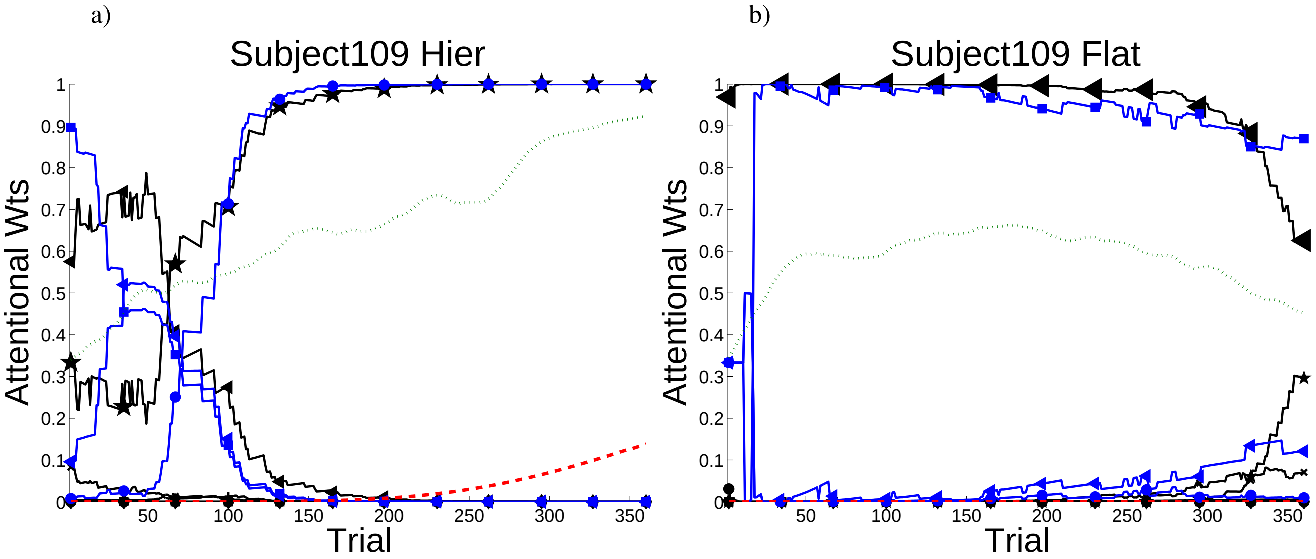

Figure 1: **a)** Full set of attentional weights for a subject who had stronger attention to the Flat conjunctive expert in the hierarchical condition (signified by asterisks, as in main text). Although there is evidence for the correct hierarchical rule (blue circles, $W_{OS|C}$), attention to overall hierarchical expert (red dashed line, W_H) does not rise until later in the block, and only minimally so. **b)** Attentional weights in the flat condition. This subject performed the flat block first, and due to a relatively low prior for the conjunctive OSC expert, the weight for this expert did not increase until late in the block. This subject's choices were best characterized by a unidimensional orientation rule (black triangles). Nevertheless, because posteriors at the end of this block were carried over to the next block (see main text) such that this subject appeared to learn conjunctively in the hierarchical block (panel a).

Neural model implementational details

The model is implemented using the Leabra framework (O'Reilly & Munakata, 2000), with the emergent neural simulation software (Aisa, Mingus, & O'Reilly, 2008), simulating the anatomical projections and physiological properties of the BG circuitry in learning, working memory and decision making (Frank, 2005; O'Reilly & Frank, 2006). Leabra uses point neurons with excitatory, inhibitory, and leak conductances contributing to an integrated membrane potential, which is then thresholded and transformed to produce a rate code output communicated to other units. In the BG model, discrete spiking can also be used, and produces similar results for decision making (but requires additional considerations to function in learning environments).

The membrane potential V_m is updated as a function of ionic conductances g with reversal (driving) potentials E according to the following differential equation:

$$\begin{aligned} C_m \frac{dV_m}{dt} &= g_e(t)\bar{g}_e(E_e - V_m) + \\ &\quad g_i(t)\bar{g}_i(E_i - V_m) + \\ &\quad g_l(t)\bar{g}_l(E_l - V_m) + \\ &\quad ..., \qquad (1) \\ & \qquad (2) \end{aligned}$$

where C_m is the membrane capacitance and determines the time constant with which the voltage can change, and subscripts e, l and i refer to excitatory, leak, and inhibitory channels respectively (and "..." refers to the possibility of adding other channels implementing neural accommodation and hysteresis). Following electrophysiological convention, the overall conductance for each channel c is decomposed into a time-varying component $g_c(t)$ computed as a function of the dynamic state of the network, and a constant $\overline{g_c}$ that controls the relative influence of the different conductances. The equilibrium potential can be written in a simplified form by setting the excitatory driving potential (E_e) to 1 and the leak and inhibitory driving potentials (E_l and E_i) of 0:

$$V_m^\infty = \frac{g_e \overline{g_e}}{g_e \overline{g_e} + g_l \overline{g_l} + g_i \overline{g_i}} \tag{3}$$

which shows that the neuron is computing a balance between excitation and the opposing forces of leak and inhibition. This equilibrium form of the equation can be understood in terms of a Bayesian decision making framework, whereby the neuron evaluates whether the excitatory evidence for the "hypothesis" it is detecting according to its synaptic weights is sufficiently greater than the evidence against that hypothesis. The excitatory net input/conductance $g_e(t)$ is computed as the proportion of open excitatory channels as a function of sending activations times the weight values:

$$g_e(t) = \langle x_i w_{ij} \rangle = \frac{1}{n} \sum_i x_i w_{ij} \tag{4}$$

The inhibitory conductance is computed as described in the next section, and leak is a constant.

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

$$y_j(t) = \frac{1}{\left(1 + \frac{1}{\gamma [V_m(t) - \Theta]_+}\right)} \tag{5}$$

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

$$y_j^*(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/(2\sigma^2)} y_j(z - x) dz \tag{6}$$

where x represents the $[V_m(t) - \Theta]_+$ value, and $y_j^*(x)$ is the noise-convolved activation for that value.

Inhibition Within Layers

For within layer lateral inhibition, Leabra uses a kWTA (k-Winners-Take-All) function to achieve inhibitory competition among units within each layer (area). The kWTA function computes a uniform level of inhibitory current for all units in the layer, such that the $k + 1$th most excited unit within a layer is generally below its firing threshold, while the kth is typically above threshold. Activation dynamics similar to those produced by the kWTA function have been shown to result from simulated inhibitory interneurons that project both feedforward and feedback inhibition (O'Reilly & Munakata, 2000), and indeed other versions of the BG model use explicit populations of striatal inhibitory interneurons, in addition to inhibitory projections from striatum to GPi/GPe, etc (e.g., Wiecki, Riedinger, von Ameln-Mayerhofer, Schmidt, & Frank, 2009). Thus, the kWTA function provides a computationally effective and efficient approximation to biologically plausible inhibitory dynamics.

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

$$g_i = g_{k+1}^{\Theta} + q(g_k^{\Theta} - g_{k+1}^{\Theta}) \tag{7}$$

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

$$g_i^{\Theta} = \frac{g_e^* \bar{g}_e (E_e - \Theta) + g_l \bar{g}_l (E_l - \Theta)}{\Theta - E_i} \tag{8}$$

where g_e^* is the excitatory net input.

Two versions of kWTA functions are typically used in Leabra. In the kWTA function used in the Striatum, g_k^{Θ} and g_{k+1}^{Θ} are set to the threshold inhibition value for the kth and $k + 1$th most excited units, respectively. Thus, the inhibition is placed to allow k units to be above threshold, and the remainder below threshold.

Other layers use the *average-based* kWTA version, where g_k^{Θ} is the average g_i^{Θ} value for the top k most excited units, and g_{k+1}^{Θ} is the average of g_i^{Θ} for the remaining $n - k$ units. This version allows for more flexibility in the actual number of units active depending on the nature of the activation distribution in the layer and the value of the q parameter (which is set to default value of .6). This flexibility is generally used for units to have differential levels of activity during settling.

Connectivity

The connectivity of the BG network is critical, and is thus summarized here (see Frank, 2006 and O'Reilly & Frank, 2006 for details and references). Unless stated otherwise, projections depicted in Figure 1 are fully connected (that is all units from the source region target the destination region, with a randomly initialized synaptic weight matrix). However the units in PFC, Striatum, SNrThal are all organized with columnar structure. Units in the first stripe of PFC represent one set of representations and project to a single column of Go and NoGo units in the Striatum, which in turn project to the corresponding column in SNrThal. Each Thalamic unit is reciprocally connected with the associated column in PFC. This connectivity is similar to that described by anatomical studies, in which the same cortical region that projects to the striatum is modulated by the output through the BG circuitry and Thalamus.

Dopamine units in the SNc project to the entire Striatum, but with different projections to encode the effects of D1 receptors in Go neurons and D2 receptors in NoGo neurons. With increased dopamine, Go units are excited while NoGo units are inhibited, and vice-versa with lowered dopamine levels. However the particular set of units that are impacted by dopamine is determined by those receiving excitatory input from sensory cortex and PFC. Thus dopamine modulates this activity, thereby affecting the relative balance of Go vs NoGo activity in those units activated by cortex.

Learning

For learning, Leabra uses a combination of error-driven and Hebbian learning. The error-driven component is the symmetric midpoint version of the GeneRec algorithm (O'Reilly, 1996), which is functionally equivalent to the deterministic Boltzmann machine and contrastive Hebbian learning (CHL), computing a simple difference of a pre and postsynaptic activation product across these two phases. For Hebbian learning, Leabra uses essentially the same learning rule used in competitive learning or mixtures-of-Gaussians which can be seen as a variant of the Oja normalization (Oja, 1983). The error-driven and Hebbian learning components are combined additively at each connection to produce a net weight change.

The equation for the Hebbian weight change is:

$$\Delta_{hebb} w_{ij} = x_i^+ y_j^+ - y_j^+ w_{ij} = y_j^+ (x_i^+ - w_{ij}) \tag{9}$$

and for error-driven learning using CHL:

$$\Delta_{err} w_{ij} = (x_i^+ y_j^+) - (x_i^- y_j^-) \tag{10}$$

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

$$\Delta_{sberr} w_{ij} = [\Delta_{err}]_+ (1 - w_{ij}) + [\Delta_{err}]_- w_{ij} \tag{11}$$

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

$$\Delta w_{ij} = \epsilon [k_{hebb} (\Delta_{hebb}) + (1 - k_{hebb})(\Delta_{sberr})] \tag{12}$$

Striatal Learning Function

Synaptic connection weights in striatal units were trained using a reinforcement learning version of Leabra. The learning algorithm involves two phases, and is more biologically plausible than standard error backpropagation. In the *minus phase*, the network settles into activity states based on input stimuli and its synaptic weights, ultimately "choosing" a response. In the *plus phase*, the network resettles in the same manner, with the only difference being a change in simulated dopamine: an increase of SNc unit firing for positive reward prediction errors, and a decrease for negative prediction errors (Frank, 2005; O'Reilly & Frank, 2006).

For the large-scale BG-PFC models used here and in O'Reilly and Frank (2006) some abstractions are used. Each stripe (group of units) in the Striatum layer is divided into Go vs. NoGo in an alternating fashion. The DA input from the SNc modulates these unit activations in the update phase by providing extra excitatory current to Go and extra inhibitory current to the NoGo units in proportion to the positive magnitude of the DA signal, and vice-versa for negative DA magnitude. This reflects the opposing influences of DA on these neurons (Frank, 2005; Gerfen, 2001; Shen, Flajolet, Greengard, & Surmeier, 2008). The update phase DA signal reflects the critic system's evaluation of the PFC updates produced by gating signals – that is, if the PFC state is predictive of reward, the striatal units will be reinforced. Learning on weights into the Go/NoGo units is based on the activation delta between the update ($++$) and plus phases:

$$\Delta w_i = \epsilon x_i (y^{++} - y^+) \tag{13}$$

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

$$g_e = \gamma [da]_+ y^+ + (1 - \gamma)[da]_+ \tag{14}$$

where $0 < \gamma < 1$ controls the degree of contrast enhancement (.5 is used in all simulations), $[da]_+$ is the positive magnitude of the DA signal (0 if negative), $y+$ is the plus-phase unit activation, and g_e is the extra excitatory current produced by the da (for Go units). A similar equation is used for extra inhibition (g_i) from negative da ($[da]_-$) for Go units, and vice-versa for NoGo units.

Dopamine and prediction errors in the "critic"

As mentioned in the main text, we used a simplified version of the critic in these simulations because they do not depend on the differences between different algorithms (e.g, temproral difference learning or "PVLV", the algorithm used in our other BG-PFC networks (O'Reilly & Frank, 2006; O'Reilly, Frank, Hazy, & Watz, 2007). To this end we include only the "PV", or primary value, part of the PVLV algorithm, which is equivalent to the standard Rescorla-Wagner delta rule, but implemented using two Leabra layers: one that represents the actual reward and is excitatory on DA signaling (PVe), and another that predicts reward and is inhibitory on DA signaling (PVi).

The PVe layer does not learn, and is always just clamped to reflect any received reward value (r). Ee use a value of 0 to reflect negative feedback, .5 for no feedback, and 1 for positive feedback (the scale is arbitrary). The PVi layer units (y_j) are trained at every point in time to produce an expectation for the amount of reward that will be received at that time. In the minus phase of a given trial, the units settle to a distributed value representation based on sensory inputs. This results in unit activations y_j^-, and an overall weighted average value across these units denoted PV_i. In the plus phase, the unit activations (y_j^+) are clamped to represent the actual reward r (a.k.a., PV_e). The weights (w_{ij}) into each PVi unit from sending units with plus-phase activations x_i^+, are updated using the delta rule between the two phases of PVi unit activation states:

$$\Delta w_{ij} = \epsilon(y_j^+ - y_j^-)x_i^+ \tag{15}$$

This is equivalent to saying that the US/reward drives a pattern of activation over the PVi units, which then learn to activate this pattern based on sensory inputs.

The distributed value representations drive the dopamine layer (VTA/SNc) activations in terms of the difference between the excitatory and inhibitory terms:

$$\delta = PV_e - PV_i \tag{16}$$

SNrThal Units

The SNrThal units provide a simplified version of the SNr/GPe/Thalamus layers abstracted from the full circuitry implemented in more basic versions of the BG circuit (e.g., Frank, 2006). They receive a net input that reflects the normalized Go - NoGo activations in the corresponding Striatum stripe:

$$\alpha_j = \left[\frac{\sum Go - \sum NoGo}{\sum Go + \sum NoGo}\right]_+ \tag{17}$$

(where $[]_+$ indicates that only the positive part is taken; when there is more NoGo than Go, the net input is 0). This net input then drives standard Leabra point neuron activation dynamics, with kWTA inhibitory competition dynamics that cause stripes to compete to update the PFC. This dynamic is consistent with the notion that competition/selection takes place primarily in the smaller GP/SNr areas, and not much in the much larger striatum (e.g., Mink, 1996; Jaeger, Kita, & Wilson, 1995). The resulting SNrThal activation then provides the gating update signal to the PFC: if the corresponding SNrThal unit is active (above a minimum threshold; .1), then active maintenance currents in the PFC are toggled.

This SNrThal activation also multiplies the per-stripe DA signal from the SNc:

$$\delta_j = snr_j \delta \tag{18}$$

where snr_j is the snr unit's activation for stripe j, and δ is the global DA signal. This implements the stripe-wise credit assignment mechanism referred to in the main text, whereby stripes that caused a Go signal receive extra modulation of phasic DA than those that did not influence updating.

PFC Maintenance

PFC active maintenance is supported in part by excitatory ionic conductances that are toggled by Go firing from the SNrThal layers. This is implemented with an extra excitatory ion channel in the basic V_m update equation (2). This channel has a conductance value of .5 when active. See Frank, Loughry, & O'Reilly, 2001 for further discussion of this kind of maintenance mechanism, which has been proposed by several researchers e.g., Lewis & O'Donnell, 2000; Lisman, Fellous, & Wang, 1999; Wang, 1999; Dilmore, Gutkin, & Ermentrout, 1999; Gorelova & Yang, 2000; Durstewitz, Seamans, & Sejnowski, 2000. The first opportunity to toggle PFC maintenance occurs at the end of the first plus phase, and then again at the end of the second plus phase (third phase of settling). Thus, a complete update can be triggered by two Go's in a row, and it is almost always the case that if a Go fires the first time, it will fire the next, because Striatum firing is primarily driven by sensory inputs, which remain constant.

Measuring prePMD Activity in Hierarchical and Flat Conditions

In the main text we reported differences in prePMD activity in Hierarchical and Flat conditions. Here we describe in more detail how this activation was recorded. Recall that the network includes both striatal input-gating and output-gating layers. The output-gating function is more relevant here because it determines the extent to which the corresponding PFC layer influences activity downstream in the network. If it is not helpful in the task to rely on activity in the PFC layer in question, the striatal layer learn to not output-gate this activity. We thus measured activity in the attentional output layer of prePMD. (In principle the same could be done in the maintenance layer, however the parameters used in these networks is such that there is always one representation in a given stripe, whether it is due to previous gating / maintenance, or else to current sensory input). Because localist units are used in these simulations, with a K value of 1 per stripe, we simply recorded the normalized firing rate of the most active unit in the layer. Simulation results showed that this activity starts relatively high in both hierarchical and flat conditions, but declines in the flat condition as NoGo activity accumulates in the striatal output-gating layer. This occurs because negative prediction errors are generated in response to any activity in the prePMD attentional layer, because this activity serves to constrain attention to only one of the other two features (orientation and shape) and this is detrimental in the flat condition. Thus the critic punishes striatal output-gating units interacting with the prePMD layer, and in turn, the activation declines. This prediction was tested in the fMRI data using the MoE to estimate prediction errors modulated by attention to hierarchical rule in individual subjects.

Bayesian update for mixture of experts model

In the main text we describe how each expert of the MoE model represents the probability that a given response will be rewarding, given the stimulus, or the probability that a given expert is rewarding (at the level of attentional weights). In each case, these probabilities are modeled as beta distributions.

The probability density function of the beta distribution is as follows:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 z^{\alpha-1}(1-z)^{\beta-1}dz}$$

where the integral in the denominator is the beta function $B(\alpha,\beta)$ and is a normalization factor that ensures that the area under the density function is always 1. The defining parameters of the posterior distribution for each response (and each expert at the level of attentional weights) are calculated after each outcome using Bayes' rule:

$$P(\alpha,\beta|\delta_1...\delta_n) = \frac{P(\delta_1...\delta_n|\alpha,\beta)P(\alpha,\beta)}{\int\int P(\delta_1...\delta_n|\alpha,\beta)d\alpha d\beta} = \frac{P(\delta_1...\delta_n|\alpha,\beta)P(\alpha,\beta)}{P(\delta_1...\delta_n)}.$$

Due to the conjugate prior relationship between binomial and beta distributions, this update is trivial without having to directly compute Bayes' equation above. The α and β parameters are updated by simply incrementing the prior α and β hyperparameters after each instance of a positive or negative prediction error δt, respectively.

$$\alpha(t+1) = \begin{cases} \alpha(t)+1 & \text{if } \delta t > 0 \\ \alpha(t) & \text{otherwise,} \end{cases} \tag{19}$$

$$\beta(t+1) = \begin{cases} \beta(t)+1 & \text{if } \delta t < 0 \\ \beta(t) & \text{otherwise,} \end{cases} \tag{20}$$

References

Aisa, B., Mingus, B., & O'Reilly, R. (2008). The emergent neural modeling system. *Neural networks*, *21*(8), 1146–1152.

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

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

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated parkinsonism. *Journal of Cognitive Neuroscience*, *17*, 51–72.

Frank, M. J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural networks : the official journal of the International Neural Network Society*, *19*, 1120–1136.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience*, *1*, 137–160.

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

Gorelova, N. A., & Yang, C. R. (2000). Dopamine d1/d5 receptor activation modulates a persistent sodium current in rats prefrontal cortical neurons in vitro. *Journal of Neurophysiology*, *84*, 75.

Hernandez-Lopez, S., Bargas, J., Surmeier, D. J., Reyes, A., & Galarraga, E. (1997). D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an l-type ca2+ conductance. *Journal of Neuroscience*, *17*, 3334–42.

Jaeger, D., Kita, H., & Wilson, C. J. (1995). Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *Journal of neurophysiology*, *72*, 2555–2558.

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

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

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

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

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

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, *8*(5), 895–938.

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

O'Reilly, R. C., Frank, M. J., Hazy, T. E., & Watz, B. (2007). Pvlv: The primary value and learned value pavlovian learning algorithm. *Behavioral Neuroscience*, *121*, 31–49.

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

Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science (New York, N.Y.)*, *321*(5890), 848–851.

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

Wiecki, T. V., Riedinger, K., von Ameln-Mayerhofer, A., Schmidt, W. J., & Frank, M. J. (2009). A neurocomputational account of catalepsy sensitization induced by d2 receptor blockade in rats: context dependency, extinction, and renewal. *Psychopharmacology*, *204*, 1–13.