Model of the Interactions between Neuromodulators and Prefrontal Cortex during a Resource Allocation Task

Suhas E. Chelian, Nicholas Oros, Andrew Zaldivar, Jeffrey L. Krichmar, Rajan Bhattacharyya

***Abstract*—Neuromodulators such as dopamine (DA), serotonin (5-HT), and acetylcholine (ACh) are crucial to the representations of reward, cost, and attention respectively. Recent experiments suggest that the reward and cost of actions are also partially represented in orbitofrontal and anterior cingulate cortices in that order. Previous models of action selection with neuromodulatory systems have not extensively considered prefrontal contributions to action selection. Here, we extend these models of action selection to include prefrontal structures in a resource allocation task. The model adapts to its environment, modulating its aggressiveness based on its successes. Selective lesions demonstrate how neuromodulatory and prefrontal areas drive learning and performance of strategy selection.**

***Index Terms*—Computational Neuroscience, Decision-Making, Policy Search, Neuromodulation, Reward, Cost**

I. Introduction

NEUROMODULATORS such as dopamine (DA), serotonin (5-HT), and acetylcholine (ACh) affect both short and long term dynamics of neural circuits that represent reward, cost, and attention in that order [1].

The ventral tegmental area (VTA) and the substantia nigra (SN) are the source of dopamine, which is related to expected reward, and incentive salience or "wanting" [2, 3]. The raphe nucleus (RN) is the source of serotonin, which is linked to cognitive control of stress, social interactions, and risk taking behavior [4, 5].

The basal forebrain is the source of acetylcholine, and appears to modulate attention and optimize information processing. Experiments conducted by Chiba et al. [6] and Baxter et al. [7] showed that the basal forebrain has specific and separate pathways for decrementing and incrementing attention: 1) ACh projections from the medial septum/vertical limb of the diagonal band (BF.MS) to the hippocampus and medial prefrontal cortex were crucial to reduce attention to irrelevant stimuli (medial prefrontal cortex includes anterior cingulate cortex, ACC), and 2) ACh projections from the substantia innominata/nucleus basalis region (BF.SI) to the amygdala and neocortex were necessary to increase attention to relevant stimuli. Ross et al. [8] have also shown that depletion of ACh in orbitofrontal cortex (OFC) impairs associative learning tasks.

Recent experiments suggest that the reward and cost of actions are also partially represented in OFC and ACC respectively (e.g., [9, 10]). Rudebeck et al., for example, trained rats to choose maze arms that yielded more food pellets either after a delay or after scaling a barrier. In the first case, a rat with an impaired ability to differentiate between reward magnitudes would be more likely to choose the lower (immediate) reward than the higher (deferred) reward. Such behavior was demonstrated with OFC lesions. ACC lesions, on the other hand, caused rats to more often pick lower (less effortful) rewards than higher (more effortful) rewards.

To better understand interactions between neuromodulatory systems and prefrontal cortical areas, we extended previous models of neuromodulatory influences on action selection [11, 12] to a resource allocation task. Based on its successes, the model modulates its aggressiveness to the environment. Learning and performance in strategy selection from neuromodulatory and prefrontal cortical areas were also affected by selective lesioning.

Manuscript received June 15, 2012. Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract number D10PC20021. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained hereon are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the U.S. Government.

S. E. Chelian and R. Bhattacharyya are with HRL Laboratories, LLC, Information Systems and Sciences Lab, Malibu, CA 90265 USA (310-317-5759; e-mail: {sechelian, rajan}@hrl.com).

N. Oros, A. Zaldivar, and J. L. Krichmar are with the Department of Cognitive Sciences at the University of California, Irvine, Irvine, CA 92697 USA (949-824-5888; email: {noros, azaldiva, jkrichma}@uci.edu)

II. Methods

A. Resource Allocation Task

Simulated agents played a variant of the multi-arm bandit game with betting [13] adapted to a resource allocation task. The task simulated a military planner's decisions to defend against an attack. Specifically, given attack probability estimates for four groups, agents must assign troops to defend against these groups. Thus, each trial "bets" troop units across four bandit arms.

Intuitively, one would assign more troops to groups that are

more likely to attack. This is known as probability matching. For example, if the agent believes groups A, B, C, and D are likely to attack with probabilities of 80%, 10%, 5%, and 5% respectively, probability matching (PM) would assign 100 troops in the following manner: 80 to group A, 10 to group B, and 5 to both group C and D (assuming the marginal utility of each troop unit is constant). A more aggressive strategy, call it PM+, would assign more resources to groups which are more likely to attack, e.g., a troop allocation of 90, 10, 0, and 0 given the previous group attack probability estimates. Conversely, a more conservative strategy, call it PM-, would reserve troops to groups which are less likely to attack; e.g., a troop allocation of 70, 14, 8, 8 for the same group attack probability estimates.

After assigning resources to the four groups, agents were directly rewarded by the number of troops they assigned to the actual attack perpetrator. For example, if group A was the true attacker, than PM+ (PM-) would be rewarded with a score of 90 (70). The cost would be the number of troops allocated to groups other than A: 10 for PM+, and 30 for PM- (14+8+8). However, if group B was the attacker, the reward of PM- exceeds that of PM+ (14 > 10). Experimental evidence suggests that agents will change their strategy based on their success rate [14, 15, 16]. Therefore, across several trials, agents should choose a more (less) aggressive strategy when estimates about group attacks are accurate (inaccurate).

B. Neural Agent

Motivated by the work of Krichmar et al. [11, 12], we created a neural model with three areas: Input, Action, and Neuromodulatory systems (Fig. 1).

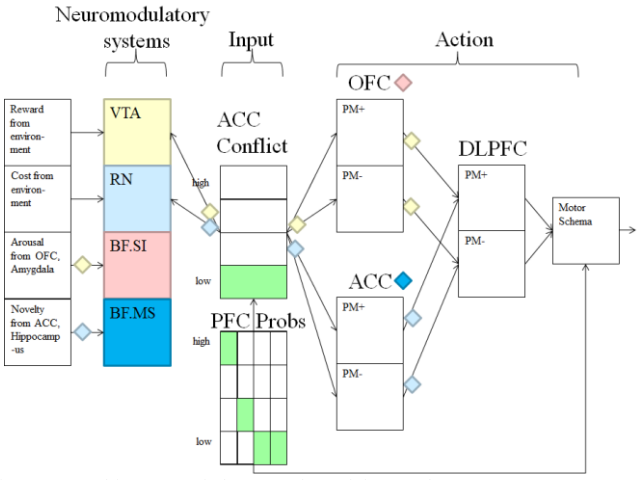

Fig. 1 – Architecture of the neural model. In the Input area, PFC Probs registers group attack probability estimates of (e.g.,) 80%, 10%, 5%, and 5%. ACC Conflict registers relatively low conflict. In the Action area, OFC and ACC neurons, each with different degrees of aggressiveness, encode the different strategies. OFC and ACC then project to DLPFC, where a motor schema takes the selected aggressiveness parameter and transforms PFC Probs into a resource allocation. In the Neuromodulatory systems area, VTA, RN, and BF.SI and BF.MS represent reward, cost, and incremental and decremental attentional pathways in that order. VTA (RN) is used to register the difference between actual and predicted reward (cost). Reward was equal to the number of troops assigned to the attacking group, while cost was equal to the number of troops used to defend against other groups. BF.SI and BF.MS control sharpening and flattening dynamics in OFC and ACC respectively. Diamonds represent corresponding neuromodulatory effects. For example, weights between ACC Conflict and OFC used the dopamine signal from VTA. Table I contains further details.

Neurons in the PFC region of the Input area held group attack probabilities estimates (PFC Probs), while neurons in the ACC region of the Input area had a localist representation for the spread of these estimates (ACC Conflict). For simplicity, we programmatically calculated a measure of dispersion on PFC Probs and placed its value into ACC Conflict. ACC has long been implicated with several performance monitoring functions including measuring conflict between different actions or outcomes (see [17] for a review). Here, however, we measure the conflict of beliefs, not actions—the conflict of group attack probability estimates, not the choice between sharpening or flattening group attack probabilities into a resource allocation. If agents had to make a binary choice for resource allocation as in typical bandit games, this conflict of beliefs would become a conflict of actions. Typically, conflict is measured with Hopfield energy (e.g., [18]). However, energy does not differentiate between flat (.25, .25, .25, .25) and peaky (e.g., 1, 0, 0, 0) probability distributions; in both cases it is 1. We chose normalized entropy instead because it differentiates between these two extremes; it measures 1 and 0 bits in that order for these distributions. Equation 1 gives the formula for normalized entropy as represented in ACC Conflict:

$$h = -\frac{1}{log_2 N}\sum_i p_i log_2 p_i \quad (1)$$

where p_i is the group attack probability estimate for group i within PFC Probs and N is the total number of groups.

There were two possible strategies in the Action area, one for PM+ and one for PM-. Each strategy had a corresponding OFC and ACC component, which then compete and project to a final decision layer perhaps in dorsolateral PFC (DLPFC), an area of PFC known to be involved in executive control of actions (e.g., [19], [20]). The site of this competition may be the dorsal medial striatum and associated components within the basal ganglia [21]. In this case, we assume OFC (ACC) units project to the direct or Go (indirect or No Go) pathway. The chosen strategy in DLPFC takes group attack probability estimates and implements a motor schema to sharpen or flatten this distribution in assigning resources. We used the Power rule to implement this function:

$$y_i = x_i{}^{\alpha} / \sum_i x_i{}^{\alpha}. \quad (2)$$

where y_i are troops assigned and x_i are group attack probability estimates both for group i and α is a parameter. When α is greater than (less than) 1, x_i's are sharpened (flattened) implementing PM+ (PM-). (As α approaches infinity, the Power rule implements winner take all competition. Flat distributions are returned when α approaches zero.) In the Action area, the chosen strategy in DLPFC is a product of long term weights in the Input area as well as short term effects from the Neuromodulatory area.

In the Neuromodulatory area, VTA, RN, BF.SI and BF.MS were simulated. The activity of these neurons was based on synaptic input from the Input and Action areas.

The synaptic connectivity of the network is shown in Fig. 1 and Table I. As in the work of Krichmar et al. [11, 12], VTA

(RN) was used to drive learning according to difference between actual and expected rewards (costs). Reward was equal to the percentage of troops assigned to the attacking group, while cost was equal to the percentage of troops used to defend against other groups. For example, given a troop allocation of 90, 10, 0, and 0 with the actual attacker being the first group, reward would be 0.9, while cost would be 0.1. BF.SI was driven by an arousal signal from OFC (possibly via the amygdala) which in turn incremented attention, or sharpened, OFC neurons. BF.MS was driven by a novelty signal from ACC (possibly via the hippocampus) which in turn decremented attention, or flattened, ACC neurons.

TABLE 1 – SYNAPTIC CONNECTIONS BETWEEN NEURAL AREAS

Row	From	To	Plasticity	Phasic neuromodulation
Input				
1	ACC Con-flict	OFC	Reward-VTA	N
2		ACC	Cost-RN	N
3		VTA	Reward-VTA	N
4		RN	Cost-RN	N
Action				
5	OFC	DLPFC	Reward-VTA	N
6		OFC^{C}	N/A	Y, BF.SI sharpens
7		BF.SI	Reward-VTA^{STP}	N
8	ACC	DLPFC^{I}	Cost-RN	N
9		ACC^{C}	N/A	Y, BF.MS flattens
10		BF.MS	Cost-RN^{STP}	N

In the To column, a ^{C} represents lateral competition as implemented by the Power rule, while an ^{I} represents inhibition. In the Plasticity column, *R* in (6) is given for long term plasticity. Reward and cost are provided from the external environment. VTA (RN) refers to the activity of the VTA (RN) neuron. For short-term plasticity, *e* in (4) is fixed at 1 except where ^{STP} appears. In those cases, (5) creates a pre-synaptic efficacy term to capture short term dynamics.

The neural activity was simulated by a mean firing rate neuron model:

$$s_i(\mathrm{t}) = \rho_i s_i(\mathrm{t}-1) + (1-\rho_i)\left(\frac{1}{1+\exp(-I_i(\mathrm{t})/\tau)}\right) \quad (3)$$

where s_i is a neuron's activation level at time t, ρ_i is a parameter that represents the persistence of a neuron, I_i, its input, and τ a parameter that controls the steepness of its activation function. The input to a neuron was the inner product of pre-synaptic connection strengths and neural activity, multiplied by an optional pre-synaptic efficacy term (e_j, defined below):

$$I_i = \sum_j e_j\, w_{ji}(\mathrm{t}-1) s_j(\mathrm{t}) \quad (4)$$

where w_{ji} was the synaptic weight from neuron j to i and s_j is the pre-synaptic neuron's activation level at time t. e_j was fixed at 1 except in the case of weights between OFC, ACC and their corresponding BF regions to capture exploitation/exploration dynamics based on gain and loss. In those cases, e_j used the short-term plasticity model of [22]:

$$\Delta e_j = r[1\text{-}e_j(\mathrm{t})]\text{-}d s_j(\mathrm{t}) e_j(t) \quad (5)$$

where e_j represents synaptic efficacy at time t, *r* and *d* dictate recovery and depletion rates in that order. When the BF.SI (BF.MS) neuron's activation exceeded a threshold, it would trigger sharpening (flattening) in OFC (ACC). For simplicity, the Power rule (2) was used to implement sharpening and flattening in OFC and ACC and winner take all choice in DLPFC where s_i was used in place of x_i, and y_i was the new value for s_i. Finally, weight updates were performed with the following equation:

$$\Delta w_{ji} = \eta s_j(\mathrm{t}) s_i(\mathrm{t}) R \quad (6)$$

where η is a learning rate, and R, the level of reinforcement, is given in Table I (Plasticity column). The Appendix lists parameter values.

A trial consisted of random group attack probability estimates being generated and their spread being measured in the Input area. Activation from the Input layer propagated into the Action and Neuromodulatory areas. After a winning strategy was selected in the Action area, the agent output its troop allocations and received feedback. This feedback in turn updated short term (BF neuron activities) and long term (weights involving VTA and RN) dynamics.

If group attack probability estimates were accurate, reward from the environment would be higher for PM+ than for PM-. In this case, weights between ACC Conflict and the PM+ strategy in OFC should increase due to higher than expected reward (Table I, row 1). Weights between ACC Conflict and VTA would also move in the same direction (Table I, row 3). There would be the same trend for weights between PM+ in OFC and DLPFC (Table I, row 5). Weights between ACC Conflict and the PM+ strategy in ACC (Table I, row 2), ACC Conflict and RN (Table I, row 4), and PM+ in ACC and DLPFC (Table I, row 8) should decrease due to lower than expected cost. Thus, on correct trials, an aggressive strategy would increment its weights relative to a conservative one, making it more likely to be picked on the next trial. Conversely, on incorrect trials, an aggressive strategy would decrement its weights due to lower than expected reward.

Weights between OFC and BF.SI and ACC and BF.MS capture short-term dynamics. In particular, when actual reward exceeds predicted reward (i.e., when correct), OFC weights to BF.SI increase (Table I, row 7). This causes BF.SI to fire more, which in turns engages lateral inhibition within OFC, sharpening its units. This positive feedback loop creates a perseverative or exploitive effect as previously selected strategies continue to be chosen. When incorrect, ACC weights to BF.MS increase (Table I, row 10), causing BF.MS to fire more, which in turns disengages lateral inhibition within ACC, flattening its units. This leads to an explorative effect by choosing strategies that were previously considered too costly.

C. Trials, simulated agents, lesions, and scenarios

A trial consisted of a set of group attack probability estimates being generated, followed by troop allocation and feedback. Each simulated agent performed 100 consecutive trials. To simulate different subjects, 50 random initial weight sets were used. The Appendix lists parameters for this initialization process.

Selective lesions were performed on neuron populations in the Neuromodulatory and Action areas as described in Table II by setting the output of those neurons to zero.

TABLE II – LESION TABLE

Condition	OFC	ACC	BF.MS	BF.SI
Control	N	N	N	N
OFC Lesion	Y	N	N	N
ACC Lesion	N	Y	N	N
BF.MS Lesion	N	N	Y	N
BF.SI Lesion	N	N	N	Y

We also created two scenarios. In the first scenario, referred to as Accurate, the attacking group matched the highest group attack probability estimates 75% of the time while in the second scenario, referred to as Inaccurate, it only matched 25% of the time. For example, in the Accurate scenario, given group attack probabilities of 80%, 10%, 5%, 5%, the attacker would be the first group in 75% of the trials and any of the other groups in 25% of the trials. Conversely, in the Inaccurate scenario, the attacker would be the first group in 25% of the trials and any of the other groups in 75% of the trials. A successful agent should pick PM+ more often in the first scenario as group probability attack estimates predict attack perpetrators well, and choose PM- more often in the second scenario because attack estimates do not accurately predict actual attacks.

III. RESULTS

A. *Intact Model*

In general, simulated agents were able to choose the strategy that maximized (minimized) expected reward (cost). In the Accurate scenario, PM+ was used an average of 74% of the time across all trials and all simulated agents. Conversely, in the Inaccurate scenario, PM+ was used 46% of the time, a difference that is statistically significant (Kolmogorov-Smirov test, $p < .05$). These results are listed in Table III.

TABLE III – MEAN PERCENTAGE OF PM+ CHOICES ACROSS ALL TRIALS, ALL SIMULATED AGENTS

	Accurate	Inaccurate
Control	74% (37%)	46% (17%)
OFC Lesion	38% (49%) *	27% (23%) *
ACC Lesion	89% (11%) *	51% (15%)
BF.MS Lesion	67% (32%)	44% (16%)
BF.SI Lesion	66% (37%)	50% (14%)

Values in parenthesis are standard deviations. A * represents a statistically significant difference (Kolmogorov-Smirov test, $p < .05$) with respect to the control condition. For the control condition, the difference between the Accurate and Inaccurate scenario is also statistically significant.

In terms of short-term dynamics, in the Accurate (Inaccurate) scenario, BF.SI was driven by OFC, which in turn triggered OFC sharpening in an average of 36.7% (0.1%) of trials across all simulated. This is because group attack probability estimates were good predictors of attackers and hence there was no reason to shift strategies on subsequent trials. This is a preservative effect across trials. BF.MS firings were inversely correlated with BF.SI firings—an average of 0.5% (68.7%) of trials across all agents for the Accurate (Inaccurate) scenarios. This corresponds to exploration of new strategies after the simulated agents incurred losses, which was more common in the Inaccurate scenario.

Analyzing per trial choices averaged across all simulated agents showed that the models were indeed learning. For example, in the first 15 trails of the either scenario, random behavior was observed across all simulated agents—roughly 50% percent of choices were PM+. However, as trials continued in the Accurate scenario, more and more aggressive choices were made, ending in an average of 87% of all simulated agents choosing PM+ for the last 15 trials. For the Inaccurate scenario, the number of aggressive choices fell so that 46% of simulated agents choose PM+ for the last 15 trials. These results are illustrated in Fig. 2.

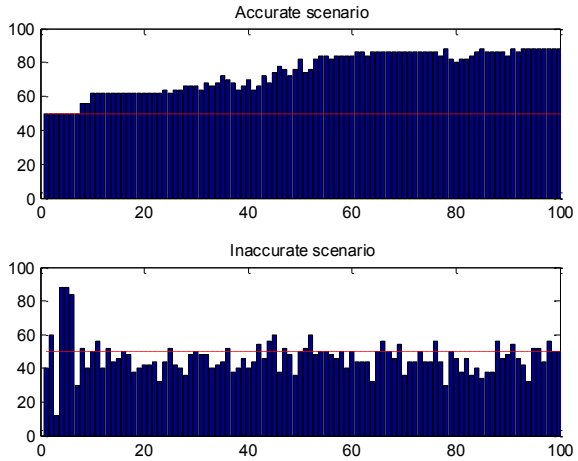

Fig. 2 – Mean percentage of simulated agents choosing PM+ for each trial in the Control condition. In both scenarios, the percentage of agents choosing PM+ in early trials was close to random (the dotted line represents 50%). However, in later trials, simulated agents were more likely to pick PM+ when group probability estimates were accurate (top) than inaccurate (bottom).

B. *Effects of Lesions*

Similar to the findings of [11, 12], we find that OFC (ACC) lesions are functionally similar to lesions of VTA (RN): OFC (ACC) lesioned simulated agents could not properly assess reward (cost). OFC lesions to simulated agents led to more conservative choices (Table III, row 3). On the other hand, ACC lesions led to more aggressive choices (Table III, row 4). For OFC lesioned agents, BF.MS firings were less common in the Accurate scenario (0.64% of trials across all simulated agents) than in the Inaccurate scenario (63.4%). As in the control condition, this demonstrates an exploration of strategies after incurring loss. (BF.SI did not fire because there was no OFC to trigger it.) For ACC lesioned agents, BF.SI firings were more common in the Accurate scenario (32.5% of trials across all simulated agents) than in the Inaccurate scenario (0.14%), which is also like the control condition. This corresponds to a preservative effect. (BF.MS did not fire because there was no ACC to trigger it.)

BF.MS lesions decreased sharpening in ACC but did not yield dramatic performance changes due to the stationary nature of each scenario. Similarly, BF.SI lesions had little

overall affect in performance. This is because both scenarios are static across trials. (A dynamic scenario, for example, would change the accuracy of group probability estimates every 20 trials.)

IV. Discussion

In this work, we presented a model that adapts to its environment, modulating its aggressiveness based on its successes. Lesions of the simulated neuromodulatory and prefrontal areas resulted in performance changes with respect to the control condition.

Just as dopamine and serotonin differentially affect reward and cost estimates, OFC and ACC perform similar roles in this task. Without OFC, simulated agents were more conservative. Without ACC, simulated agents were more aggressive.

ACh has been shown to be important to increment and decrement attention. Upregulation of attention due to exploration after loss was more common in the Inaccurate scenario than the Accurate scenario. Conversely, downregulation of attention due to exploitation after reward was more common in the Accurate scenario than the Inaccurate scenario. However, due to the static scenarios, little change in end performance was evident. Indeed, this is in agreement with the finding of Robbins and Roberts [23] that depletion of prefrontal acetylcholine impairs reversal learning, but not set formation or set shifting.

A. Comparison to Other Models

Reinforcement learning and evolutionary algorithms have also been used to model n-arm bandit and resource allocation tasks (e.g., [13, 24]). Unfortunately, as there are large number of variants for each task (e.g., binary v. analog bets, stationary v. non-stationary environments, competitors and collaborators, etc.), comparison of this model's performance with others is difficult. Our primary motivation, however, was to understand how neuromodulation affects learning and performance in prefrontal cortical areas involved in decision making. Perhaps the most comparable model to the present work is that of Litt et al. [25], which seeks to model prospect theory and decision affect theory using brain regions such as OFC, ACC, and dopaminergic and serotoninergic areas. Their model, however, does not model acetylcholinergic influences and has only been demonstrated on binary decisions.

B. Future Directions

A number of extensions to this model are possible. Cognitive phenotypes such as degree of loss or risk aversion are one avenue to explore [26]. More loss averse individuals, for example, might have a higher gain on ACC to DLPFC connections leading to more conservative choices. Similarly, more risk averse people might model the variance of a strategy's payoff and suppress more risky choices. In addition, while we have primarily explored phasic dopamine and serotonin levels, tonic components of these neuromodulators are also known to affect decision making [27]. A higher tonic dopamine level, for example, would be less sensitive to changes in reward and hence would be less likely to switch strategies. Increasing the neurofidelity of the cortico-striatal-cortical connections between OFC, ACC, and DLPFC is yet another avenue to purse. The two discrete strategies in this model could be merged into a single strategy with an analog parameter, and finally the role of noradrenaline in exploitation/exploration dynamics could also be incorporated [28].

APPENDIX

Neuron and learning rate parameters were constant for all neuron populations. Short-term plasticity was only used for specific connections as listed in Table I; otherwise its term was set to 1. Initial weights across simulated agents were drawn from a normal distribution. The number of boxes depicted in Fig. 1 were the number of neurons used for each region except for ACC Conflict; there six neurons were used.

TABLE IV – PARAMETER VALUES

Parameter	Value	Notes (e.g., equation reference)
α for PM+	2.	(2)
α for PM-	0.5	(2)
ρ_i	0.001	(3)
τ	0.25	(3)
Mean for initial weights	0.25	(4)
Standard deviation for initial weights	0.0625	(4)
r	0.05	(5)
d	0.1	(5)
η	0.25	(6)
Threshold for sharpening in OFC	0.66	N/A
Threshold for flattening in ACC	0.66	N/A
α for sharpening in OFC	2.	(2)
α for flattening in ACC	0.75	(2)
α for competition in DLPFC	100.	(2)

References

[1] J. Krichmar, "The neuromodulatory system - a framework for survival and adaptive behavior in a challenging world," *Adaptive Behavior*, vol. *16*, no. *6*, pp. 385–399, Dec. 2008.

[2] W. Schultz, "Dopamine neurons and their role in reward mechanisms," *Curr. Opin. Neurobiol.*, vol. *7*, no. *2*, pp. 191-197, April 1997.

[3] K. Berridge, "Motivation concepts in behavioral neuroscience," *Physiol. Behav.*, vol. *81*, no. *2,* pp. *179-209*, April 2004.

[4] M. Millan, "The neurobiology and control of anxious states," *Prog. Neurobiol.*, vol. *70*, no. *2*, pp. 83-244, June 2003.

[5] M. Crockett, L. Clark, G. Tabibnia, M. Lieberman, and T. Robbins, "Serotonin modulates behavioral reactions to unfairness," *Science*, vol. *320*, no. *5884*, p. 1739, June 2008.

[6] A. Chiba, D. Bucci, P. Holland, and M. Gallagher, "Basal forebrain cholinergic lesions disrupt increments but not decrements in conditioned stimulus processing," *J. Neurosci.*, vol. *15*, no. *11*, pp. *7315-7322*, Nov. 1995.

[7] M. Baxter, P. Holland, and M. Gallagher, "Disruption of decrements in conditioned stimulus processing by selective removal of hippocampal

cholinergic input," *J. Neurosci.*, vol. *17*, no. *13*, pp. *5230-5236*, July 1997.
[8] R. Ross, J. McGaughy, and H. Eichenbaum, "Acetylcholine in the orbitofrontal cortex is necessary for the acquisition of a socially transmitted food preference," *Learn. Mem.,* vol. *12*, no. *3*, pp. 302-306, May 2005.
[9] P. Rudebeck, M. Walton, A. Smyth, D. Bannerman, and M. Rushworth, "Separate neural pathways process different decision costs," *Nature Neurosci.*, vol. *9*, no. *9*, pp. *1161–1168*, Aug. 2006.
[10] M. Rushworth, T. Behrens, P. Rudebeck, and M. Walton, "Contrasting roles for cingulated and orbitofrontal cortex in decisions and social behavior," *Trends in Cog. Sci.*, vol. *11*, no. *4*, pp. 168-176, April 2007.
[11] N. Oros, A. Chiba, D. Nitz, M. Avery, and J. Krichmar, "Simulation of the neuromodulatory pathways responsible for incrementing and decrementing attention during reversal learning and set-shifting," In Proceeding of the Society for Neuroscience, Washington, D.C., 2011.
[12] A. Zaldivar, D. Asher, and J Krichmar, "Simulation of how neuromodulation influences cooperative behavior," in *Simulation of Adaptive Behavior: From Animals to Animats, Lecture Notes on Artificial Intelligence*. Berlin, Germany, Springer-Verlag, 2010, pp. 649-660.
[13] A. Niculescu-Mizil, "Multi-armed bandits with betting - on-line learning with limited feedback," in Conference on Learning Theory, Montreal, 2009.
[14] R. Cools, K. Nakamura, and N. Daw, "Serotonin and dopamine: unifying affective, activational, and decision functions," *Neuropsychopharm.*, vol. *36*, no. *1*, pp. *98-113*, Jan. 2011.
[15] F. Denk *et al.*, "Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort." *Psychopharm.* vol. *179*, no. *3*, pp. *587-596*, Dec. 2005.
[16] P. Roussos, S. Giakoumaki, S. Pavlakis, and P. Bitsios, "Planning, decision-making and the COMT rs4818 polymorphism in healthy males," *Neuropsychologia,* vol. *46*, no. *2*, pp. 757-763, Oct. 2008.
[17] W. Alexander and J. Brown, "Computational models of performance monitoring and cognitive control," *Topics in Cognitive Sci.*, vol. *2*, no. *4*, pp. *658-677*, April 2010.
[18] M. Botvinick, T. Braver, D. Barch, C. Carter, J. Cohen, "Conflict monitoring and cognitive control," *Psychological Rev.*, vol. *108*, no. *3*, pp. 624-65, July 2001.
[19] M. Sakagami, X. Pan and B. Uttl, "Behavioral inhibition and prefrontal cortex in decision-making," *Neural Networks*, vol. *19*, no. *8*, pp. 1255-1265, Sept. 2006.
[20] M. Lebedev, and S. Wise, "Insights into seeing and grasping: distinguishing the neural correlates of perception and action," *Behav. Cogn. Neurosci. Rev.,* vol. *1*, no. *2*, pp. *108–129,* June 2002.
[21] S. Haber, K. Kim, P. Mailly, R. Calzavara, "Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, proving a substrate for incentive-based learning," *J. Neurosci.*, vol. *26*, no. *32*, pp. 8368-8376, Aug. 2006.
[22] H. Markram and M. Tsodyks, "Redistribution of synaptic efficacy between neocortical pyramidal neurons," *Nature*, vol. *382*, no. *6594*, pp. 807–810, Aug. 1996.
[23] T. Robbins and A. Roberts, "Differential regulation of front-executive function by the monoamines and acetylcholine," *Cerebral Cortex,* vol. *17*, no. *supplement 1*, pp. i151-i160, Sept. 2007.
[24] J. Alcaraz and C. Maroto, "A robust genetic algorithm for resource allocation in project scheduling," *Annals of Operations Res.*, vol. *102*, no. *1*, pp. 83-109, Feb. 2001.
[25] A. Litt, C. Eliasmith, P. Thagard, "Neural affective decision theory: choices, brains, and emotions," *Cognitive Systems Res.*, vol. *9*, no. *4*, pp. 252-273, Oct. 2008.
[26] G. Loewenstein, S. Rick, and J. Cohen, "Neuroeconomics," *Annu. Rev. Psychol.*, vol. 59, no. *1*, pp. 647-672, Jan. 2008.
[27] C. Breitenstein et al., "Tonic dopaminergic stimulation impairs associative learning in healthy subjects," *Neuropsychopharm.*, vol. *31*, no. *11*, pp. 2552–2564, July 2006.
[28] G. Aston-Jones G and J. Cohen, "An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance," *Annu. Rev. Neurosci.*, vol. *28*, no. *1*, pp. 403-450, July 2008.