Simulation of How Neuromodulation Influences Cooperative Behavior

Andrew Zaldivar, Derrik E. Asher, and Jeffrey L. Krichmar

Department of Cognitive Sciences
University of California, Irvine
Irvine, CA 92697-5100
{azaldiva,dasher,jkrichma}@uci.edu

Abstract. Neuromodulators can have a strong effect on how organisms cooperate and compete for resources. To better understand the effect of neuromodulation on cooperative behavior, a computational model of the dopaminergic and serotonergic systems was constructed and tested in games of conflict and cooperation. This neural model was based on the assumptions that dopaminergic activity increases as expected reward increases, and serotonergic activity increases as the expected cost of an action increases. The neural model guided the behavior of an agent that played a series of Hawk-Dove games against an opponent. The agent adapted its behavior appropriately to changes in environmental conditions and to changes in its opponent's strategy. The neural agent tended to engage in Hawk-like behavior in low-risk situations and Dove-like behavior in high-risk situations. When the simulated dopaminergic activity was greater than the serotonergic activity, the agent tended to escalate a fight. These results suggest how the neuromodulatory systems shape decision-making and adaptive behavior in competitive and cooperative situations.

Keywords: Dopamine; Serotonin; Cooperation; Game Theory; Computational Neuroscience; Decision-Making.

1 Introduction

Neuromodulators, such as dopamine (DA) and serotonin (5-HT), are known to be important in predicting rewards, costs, and punishments. Dopamine activity (DA), which originates in the ventral tegmental area (VTA) and the substantia nigra (SN), appears to be linked to expected reward [1], and incentive salience or "wanting" [2]. Serotonin (5-HT), which originates in the Raphe nucleus, appears to be related to cognitive control of stress, social interactions, and risk taking behavior [3], [4]. The structures that are innervated by 5-HT and their connecting circuits modulate the behavioral response to threats and risks, that is, behaviors that are typically thought to reflect the anxiety state of the organism [3]. Whereas DA is related to the expected reward of a given decision, 5-HT could be related to the expected cost of a decision.

S. Doncieux et al. (Eds.): SAB 2010, LNAI 6226, pp. 649–660, 2010.

Game theory has been useful for understanding risk-taking and cooperation [5]. Of particular interest are studies in which neuromodulators were depleted or altered, while subjects play games. In one study, subjects, who were 5-HT depleted through dietary changes, cooperated less in a Prisoner's Dilemma game [6]. In an Ultimatum game study, 5-HT depleted subjects tended to reject monetary offers more than control subjects when they deemed the offers to be unfair [4]. Moreover, a recent study has shown that individuals with lower levels of dopamine in the prefrontal cortex tended to take less risks in a gambling task [7].

To better understand the roles of dopamine and serotonin during decision-making in games of conflict, we developed a computational model of neuromodulation and action-selection, based on the assumption that DA levels are related to the expected reward of an action, and 5-HT levels are related to the expected cost of an action. An agent, whose behavior was guided by the neural model, played the Hawk-Dove game, where players must choose between confrontational and cooperative tactics [5], [8]. The model makes predictions of how neuromodulatory activity can shape behavior under different environmental and competitive situations.

2 Methods

Game Playing. A *game* consisted of two agents (*Neural* and *Opponent*) taking a single action in response to a territory of interest (TOI). At the start of each game, the agents were randomly placed in a square grid (not occupying the same area) and were modeled to approach the neutral TOI at the same speed. The agent that arrived at the neutral TOI first had the opportunity to take either of the two possible actions: *Escalate* (i.e., an aggressive, confrontational tactic) or *Display* (i.e., a nonviolent, cooperative tactic). The agent that arrived second responded with one of the two aforementioned actions. After each game, payoff was calculated and plastic connections were updated. The payoff matrix for this game is given in Table 1. If both agents *Escalate*, they received a penalty that was either a serious injury (large penalty) or just a scratch (small penalty). The probability of serious injury was set to 0.25 or 0.75 at the start of the game. If both agents *Display*, they share the TOI resource. If one agent escalated and the other displayed, the agent that escalated gets the entire resource. A *series* consisted of 100 games with a given parameter set. At the start of each series, the neural network was initialized and the *Neural* agent was considered "naïve", that is, the weights of the network were set to their initial values (see next section). For each parameter set, the two agents played 100 Hawk-Dove series with a different random number seed.

Table 1. Payoff matrix for Hawk-Dove game between players A and B. V is the value of the resource and is set to 0.60. D is the damage incurred when both players escalate. D is set to 1.60 for serious injury and 0.62 for a scratch. The probability of a serious injury is 0.25 or 0.75.

	B. Escalate	**B. Display**
A. Escalate	A: (V–D)/2, B: (V–D)/2	A: V, B: 0
A. Display	A: 0, B: V	A: V/2, B: V/2

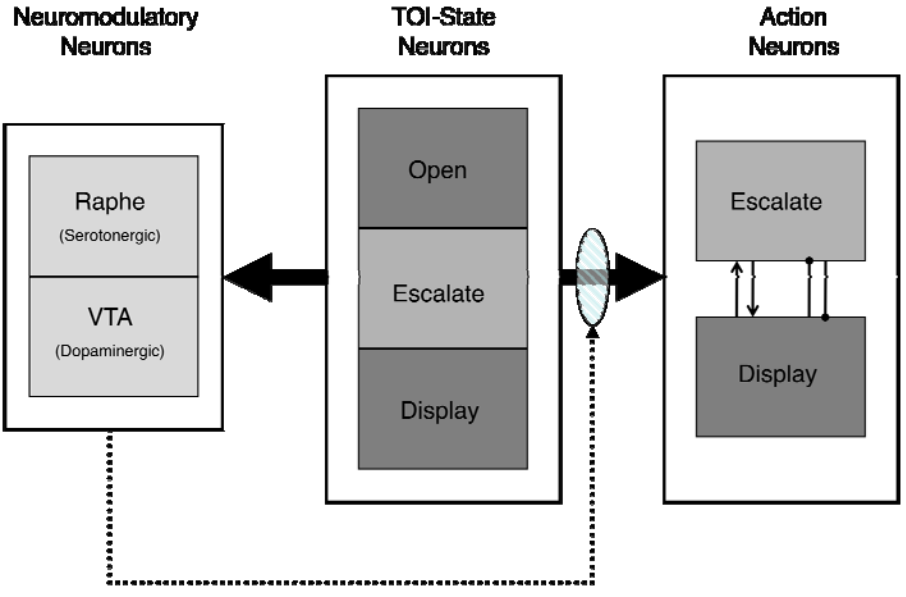

Fig. 1. The diagram shows the architecture of the neural model (two *Neuromodulatory*: Raphe and VTA; three *TOI-State*: Open, Escalate, and Display; and two *Action*: Escalate and Display). The solid arrows extending from the *TOI-State* neurons represent all-to-all connections. The thick arrows represent plastic pathways. The dotted arrows and shaded ovals represent neuromodulatory pathways. Within the *Action* neurons, the line with the arrow at the end represent excitation, and the line with the dot at the end represent inhibition.

Neural Agent. A neural network controlled the behavior of the *Neural* agent. The neural network had three areas: *TOI-State*, *Action*, and *Neuromodulatory* (Fig. 1). The *TOI-State* included three neurons that corresponded to the possible states of the TOI the *Neural* agent may observe: 1) *Open*. The *Neural* agent reached the TOI first. 2) *Escalate*. The *Opponent* agent reached the TOI first and escalated a conflict. 3) *Display*. The *Opponent* agent reached the TOI first but did not start a conflict. The equation for the activity of each of these neurons (n_i) was set based on the current state of the TOI:

$$n_i = \begin{cases} 0.75 + rnd(0.0,\ 0.25); & i = TOIState \\ rnd(0.0,\ 0.25); & Otherwise \end{cases} \quad (1)$$

where rnd(0.0,0.25) was a random number uniformly distributed between 0.0 and 0.25. The *Action* area included two neurons: 1) *Escalate*. The *Neural* agent escalated a conflict. 2) *Display*. The *Neural* agent did not start a conflict or retreated if the *Opponent* agent escalated. The neural activity was based on input from *TOI-State* neurons and neuromodulation. Lastly, the *Neuromodulatory* area included two neurons: 1) *Raphe*. A simulated raphe nucleus, which is the source of serotonergic

neuromodulation. 2) *VTA*. A simulated ventral tegmental area, which is the source of dopaminergic neuromodulation. The synaptic connectivity of the network is shown in Fig. 1 and in Table 2, and was all-to-all. Some of these connections were subject to synaptic plasticity and phasic neuromodulation, where the activity of *Neuromodulatory* neurons affected the synaptic efficacy.

Table 2. Synaptic connections between neural areas

From	To	Initial Weight	Plastic	Phasic Neuromodulation
TOI-State	*Action*	0.1	Y	Y
TOI-State	*Neuromodulatory*	0.1	Y	N
Action-Escalate	*Action-Display*	0.1	N	N
Action-Escalate	*Action-Display*	-0.1	N	Y
Action-Display	*Action-Escalate*	0.1	N	N
Action-Display	*Action-Escalate*	-0.1	N	Y

The neural activity was simulated by a mean firing rate neuron model, where the firing rate of each neuron ranged continuously from 0 (quiescent) to 1 (maximal firing). The equation for the mean firing rate neuron model was:

$$s_i(t) = \rho_i s_i(t-1) + \left(1-\rho_i\right)\left(\frac{1}{1+\exp\left(-5I_i(t)\right)}\right) \tag{2}$$

where t was the current time step, s_i was the activation level of neuron i, ρ_i was a constant set to 0.1 and denoted the persistence of the neuron, and I_i was the synaptic input. The synaptic input of the neuron was based on pre-synaptic neural activity, the connection strength of the synapse, and the amount of neuromodulator activity:

$$I_i(t) = rnd(-0.5,0.0) + \sum_j nm(t-1)w_{ij}(t-1)s_j(t-1) \tag{3}$$

where w_{ij} was the synaptic weight from neuron j to neuron i, and nm was the level of neuromodulator at synapse ij. Phasic neuromodulation had a strong effect on action selection and learning. During phasic neuromodulation, synaptic projections from sensory systems and inhibitory neurons are amplified relative to recurrent or associational connections [9]. In our model, the *TOI-State* to *Action* neurons represented sensory connections and the excitatory *Action*-to-*Action* neurons represented the associational connections. To simulate the effect of phasic neuromodulation, inhibitory and sensory connections were amplified by setting *nm* (equation 3) to ten times the combined average activity of the simulated Raphe, and VTA neurons. Otherwise, *nm* was set to 1 for recurrent or association connections. The last column of Table 2 lists connections amplified by phasic neuromodulation. In simulation studies [10] and robotic experiments [11], this mechanism was shown to be effective in making the network exploitive when neuromodulation levels were high and exploratory when neuromodulation levels were low.

Action selection depended on the summed activity of the *Action* neurons after the neural agent reached the TOI. When the *Neural* agent reached the TOI, neural activities of the *Action* and *Neuromodulator* neurons were calculated for ten time-steps

(equations 1-3). The Action neuron with the largest total activity during those ten time-steps dictated the action taken (e.g. if the total *Display* activity was greater then *Escalate*, the agent displayed).

After both the *Neural* and *Opponent* agents chose a tactic, a learning rule, which depended on the current activity of the pre-synaptic neuron, the post-synaptic neuron, the overall activity of the neuromodulatory systems and the payoff from the game, was applied to the equation for the plastic connections (see Table 2):

$$\Delta w_{ij} = \alpha * nm(t-1)s_j(t-1)\left(s_i(t-1)\right) * R \tag{4}$$

where s_j was the pre-synaptic neuron activity level, s_i was the post-synaptic neuron activity level, α was a learning rate set to 0.1, *nm* was the average activity of all neuromodulatory neurons, and *R* was the level of reinforcement based on payoff and cost (equation 5). The pre-synaptic neuron (s_j) in equation 4 was the most active *TOI-State* neuron. The post-synaptic neuron (s_i) could either be the most active *Action* neuron, the *Raphe* neuron, or the *VTA* neuron. Weights were normalized by the square root of sum of squared weights. The level of reinforcement (R, equation 4) was:

$$R = \begin{cases} (\text{Reward - } VTA) - (Cost - Raphe); & \text{TOI - State} \rightarrow \text{Action connection} \\ \text{Reward - VTA}; & \text{TOI - State} \rightarrow \text{VTA connection} \\ Cost - Raphe; & \text{TOI - State} \rightarrow \text{Raphe connection} \end{cases} \tag{5}$$

where the Reward was the *Neural* agent's payoff from Table 1 divided by the maximum possible reward. It was assumed that 5-HT plasticity was based on the predicted cost of an action and DA plasticity was based on the predicted reward of an action. If there was an error in this prediction, weights changed according to equations 4 and 5. If the Raphe or VTA accurately predicted the respective cost or payoff of an action, learning ceased. The *Neural* agent's cost was 1 if seriously injured, the ratio of scratch to serious injury (i.e., 0.3875, Table 1) if scratched, or zero otherwise. The *Neural* agent's reward was set to 1 if it won the resource, 0.5 if it split the resource, and zero otherwise.

Opponent Agent. The *Opponent* followed one of three strategies. In one strategy, referred to as the *Statistical* model, the agent had a probability of escalation independent of the *Neural* agent's tactics, which was set at the beginning of the game to 0.25 or 0.75. In the second strategy, referred to as *Tit-For-Tat* (*TT*), the computer model always repeated the *Neural* agent's previous move. The only exception to this rule was if the *Opponent* agent reached the TOI first in the opening game, in which the *Opponent* opened with a Display. *TT* is a simple, yet effective strategy in game theory, which has shown to be successful in game playing tournaments [8]. In the third strategy, referred to as *Win-Stay, Lose-Shift (WSLS)*, the *Opponent* agent would win and stay with the same action in the following situations: the *Opponent* agent's *Escalate* is met with the *Neural* agent's *Display* or the *Opponent* agent's *Display* is matched by a *Neural* agent's *Display*, otherwise the *Opponent* agent resorted to a lose and shift action [12]. As with the *TT* strategy, the *WSLS* opponent would open with a *Display* action if it arrived at the TOI first on the first game.

3 Results

Adopted Strategies. During the course of a series, the *Neural* agent learned to adopt different strategies depending on the chance of serious injury and its *Opponent's* strategy. To ensure that these strategies did not occur by chance, 100 randomly behaving agents played against all three *Opponents*. The random agents had lesions (i.e. activity set to zero) of both the simulated VTA and Raphe, which resulted in no learning occurring (equation 4). The 95% confidence interval was used as the cutoff for gauging non-random behavior in the random agents. This cutoff corresponded to the probability of selecting a particular action in response to a given *TOI-State* greater than 65% or less than 35% of the time.

The *Neural* agent adapted its behavior depending on its opponent's strategy and environmental conditions (Fig. 2). In response to a given *TOI-State*, the agent could respond randomly (i.e. within the 95% confidence), or significantly tend toward escalation or displaying. There are a total of 27 possible outcomes the *Neural* agent can take with respect to the three different states of the *TOI*. Only a few of these outcomes emerged in the simulations, and these outcomes are represented in Fig. 2 as a triplet pairing (i.e., EEE, DDE, UDE, etc.). The first value in the triplet pairing corresponds to the expected action when the *TOI-State* was *Open*. The second represents the anticipated action when the *TOI-State* was *Escalate*. The third value denotes the expected outcome when the *TOI-State* was *Display*. These triplets are associated with a color spectrum, where aggressive outcomes ('E' in the triplet) are denoted red, passive outcomes ('D' in the triplet) are denoted in blue, and values that do not fall within either outcome ('U' in the triplet) are denoted in yellow.

Against all three opponents, the *Neural* agent adopted *Hawk-like* behavior in "safe" environments, where the probability of serious injury was 0.25 (top row, Fig. 2), and *Dove-Like* behavior in "harsh" environments, where the probability of serious injury was 0.75 (bottom row, Fig. 2). Figure 2 shows an increase in the adoption of 'DDE' strategy (*Neural* agent displayed when the *TOI-State* was *Open* and *Escalate*, and escalated when the *TOI-State* was *Display*) as the probability of serious injury or an opponent escalating increased. This demonstrates that in situations where the *Neural* agent was in a competitive, antagonistic environment, the *Neural* agent tended to behave in a *Dove-like* way (displaying a large proportion of the games in a series). Conversely, Figure 2 also shows an increase in aggressive strategies (i.e., EEE, *Neural* agent escalated when the *TOI-State* was *Open*, *Escalate* and *Display*) as the probability of serious injury or an opponent escalating decreased. This illustrates that in circumstances where the *Neural* agent was in a cooperative, forgiving environment, it tended to adopt more *Hawk-like* behavior (escalating in a larger proportion of the games in a series).

Simulated lesion experiments were carried out to test the effect of neuromodulation on behavior. An intact neuromodulatory system was necessary for appropriate behavior (see Table 3). When the serotonin was removed from the system, by simulated lesions to the Raphe, the *Neural* agent's behavior became more *Hawk-like,* even when the chance of serious injury was high (*Harsh* column in Table 3). When the simulated VTA was lesioned, effectively removing dopaminergic input to the system, the *Neural* agent's behavior became more *Dove-like* (fewer escalations) in all environments.

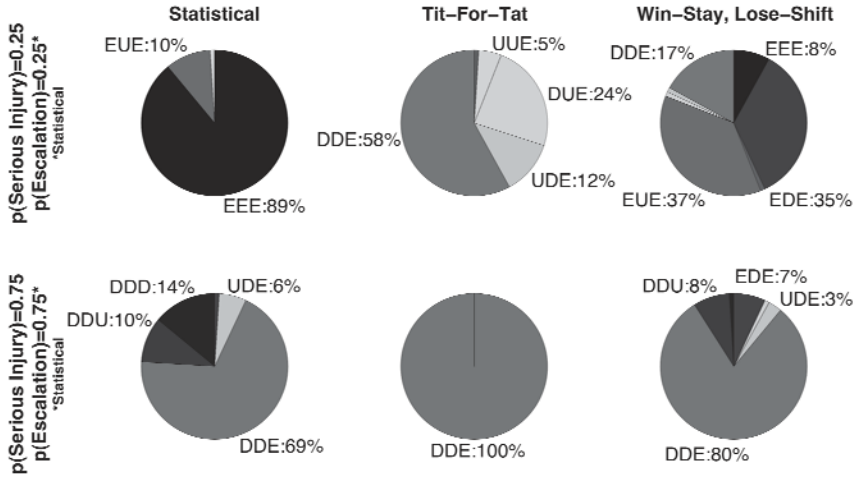

Fig. 2. The pie charts show the proportion of probable actions taken by the *Neural* agent in 100 series of games. There are three *TOI-State* areas (Open, Escalate, and Display), and three outcomes the *Neural* agent can commit to: Escalate (E), Display (D) or Undecided (U). Undecided represents random choice between 'E' and 'D'. The labels represent the *Neural* agent's response to the three *TOI-State* areas. Strategies that are *Dove-like* are displayed in blue, *Hawk-like* are displayed in red, and arbitrary strategies displayed in yellow.

Table 3. Percentage of Escalation for the *Neural* agent

	Control		**Raphe Lesion**		**VTA Lesion**	
	Safe	*Harsh*	*Safe*	*Harsh*	*Safe*	*Harsh*
Statistical	97.65%	10.00%	99.06%	92.86%	34.79%	7.14%
TT	34.15%	13.64%	81.82%	81.82%	24.74%	12.50%
WSLS	93.22%	9.09%	96.88%	96.88%	20.93%	8.22%

The *Neural* agent adapted its behavior to its *Opponent's* strategy. Against the *TT* opponent, the *Neural* agent oscillated between escalating and displaying in successive games. In essence, the *Neural* agent learned to adopt a *TT* strategy against this opponent, which yielded approximately equal reward to both agents. The oscillating neuromodulatory activity corresponded to the alternating actions taken by both agents (Fig. 3A). Against the *WSLS* opponent, the *Neural* agent created opportunities for high payoffs. The high-expected cost and reward were reflected in the serotonergic and dopaminergic activity when both agents escalated (see Fig. 3B: bottom plot, games 79, 82, or 86). In these examples, the *Neural* agent escalated first and its *Opponent* escalated second (Fig. 3B: top plot, games 79, 82, or 86). The *Neural* agent learned that this tactic caused the *Opponent* agent to 'lose-shift' towards *Display* in the following game, which could be taken advantage of by escalating (Fig. 3B: top plot, games 80, 83, or 87). This tactic resulted in a maximal reward to the *Neural*

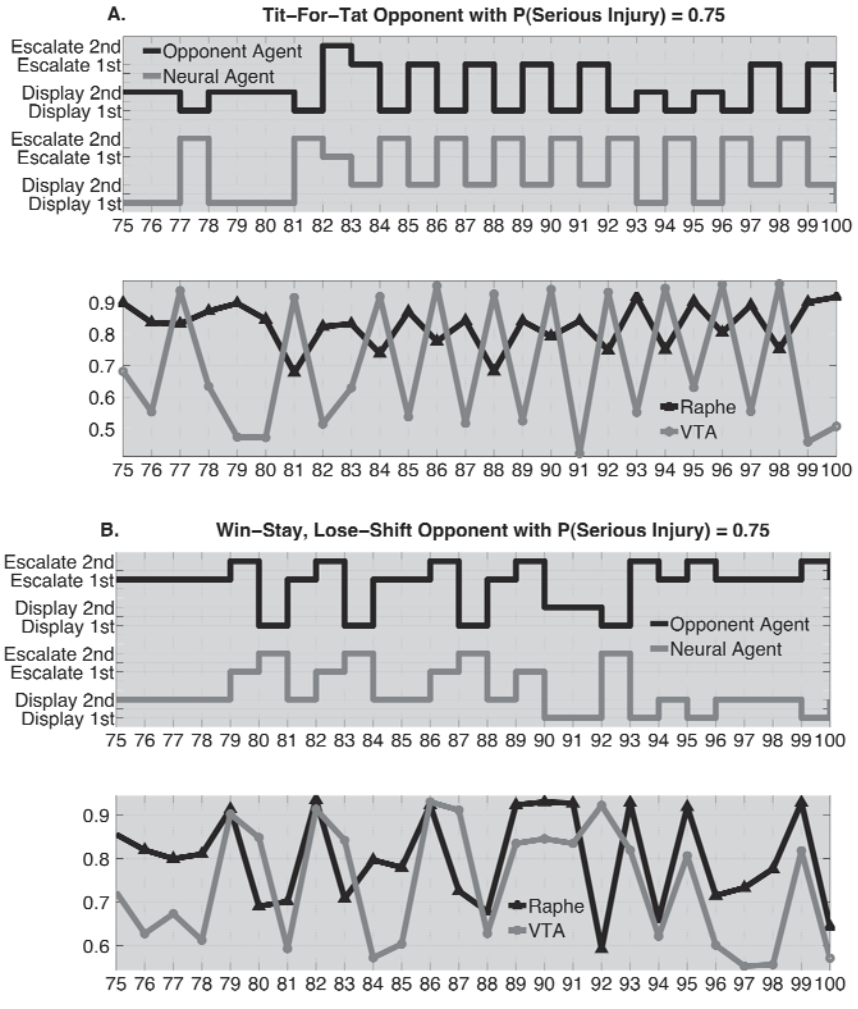

Fig. 3. Actions taken by the *Neural* and *Opponent* agents during the last 25 games of a single series, and the corresponding neuromodulatory activity for the *Neural* agent. The stair plots located on the top half of A and B, are the actions taken by both the *Neural* (green) and *Opponent* (black) agents. The line plots located in the bottom of A and B represent the neuromodulatory activity for the *Neural* agent during the same 25 games of the same series. The red line represents the Raphe activity, and the blue line represents the VTA activity. **A.** *Control* agent versus the *TT* opponent. **B.** *Control* agent versus *WSLS* opponent.

agent but caused the *Opponent* agent to 'lose-shift' back to *Escalate* in the following game (see Figure 3B: top plot, games 81, 84, or 88).

The neural response of the simulated neuromodulators appears to govern the *Neural* agent's actions (Fig. 3). When the VTA activity dropped below the Raphe activity, the neural agent displayed. That is, Raphe activity may be acting as a threshold for the expected cost of upcoming actions, whereas the VTA activity rises and falls based on

the expected reward. When the expected reward is lower than the expected cost, the *Neural* agent tended to display. For example, when a *Neural* agent behaved *Dove-like*, its serotonin activity was high relative to the dopamine activity due to the low expected reward from displaying (see Fig. 3A: games 78-80). In addition, the oscillatory actions taken by the *Neural* agent (see top Fig. 3A: games 84-99), are exactly matched by the oscillatory VTA neuromodulatory activity (see Fig. 3A: games 84-99) rising above and falling below the Raphe neuromodulatory activity. The low fluctuation in Raphe values from one game to the next in Fig. 3A result from the precision of predicted cost when playing a highly predictable opponent using the *TT* strategy. Predicted cost was not as regular for the *Neural* agent when playing against *the WSLS* opponent, which is why the Raphe neuromodulatory activity fluctuated more in Fig. 3B (bottom plot). Although the Raphe activity fluctuated more when playing against the *WSLS* opponent, the actions taken by the *Neural* agent were consistent with the neuromodulatory activity. Thus, the results from the simulated neuromodulatory activity of Fig. 3 suggest that the Raphe neural activity acts as a threshold for aggressive (escalate) or non-aggressive (display) actions taken by the *Neural* agent.

4 Discussion

In the present paper, we showed that an agent, whose behavior was guided by a computational model of the neuromodulatory system, learned to adjust its strategy appropriately depending on environmental conditions and its opponent's strategy in the Hawk-Dove game. The model makes several predictions on how the activity of neuromodulatory systems can lead to appropriate action selection in competitive and cooperative environments.

In constructing the model, it was assumed that DA activity increased as expected reward increased, and that 5-HT activity increased as the expected cost of an action increased. DA appears to be important for reward anticipation [1], and the "wanting" of things, that is, the motivation process in acquiring an object [2]. Thus, having DA activity related to payoff in a game appears to be a reasonable assumption. 5-HT activity appears to modulate behavioral response to risks, stress, threats [3], [13]and social anxiety in primates [14], all of which have a cost associated with them. Moreover, reduced 5-HT transmission is associated with a release of aversive or punishing responses [15]. These assumptions are similar to a model proposed by Daw Kakade and Dayan in which dopamine and serotonin levels track predicted rewards and punishments [16]. However, our model differs in that punishments and rewards are not necessarily mutually inhibitory. Our model takes into consideration that an action could have independent costs and rewards associated with it (i.e., an action may have a high predicted reward, and a high predicted cost).

Given these assumptions, the *Neural* agent adjusted its strategy depending on environmental conditions and on its *Opponent's* strategy (Fig. 2). For example, in situations where it was more likely to sustain a serious injury, the *Neural* agent's behavior became more *Dove-like*. Because the *Neural* agent learned that there was an increased cost and decreased reward to be expected by escalating a confrontation in these harsher conditions, it adapted its strategy to increase in *Display* actions (Fig. 2). No matter which *Opponent* the *Neural* agent faced, it learned to alter its strategy to take

advantage of a no cost escalation in response to its *Opponent's* displaying first. This can be seen in Fig. 2 for all tactics that end in 'E' (e.g., DDE or EDE).

The adaptive behavior demonstrated by the *Neural* agent required an intact neuromodulatory system in which the agent could evaluate the expected cost and the expected reward of a given action. Lowering the simulated serotonin levels resulted in *Hawk-like* tactics that were similar to uncooperative behavior seen in human studies where serotonin levels were lowered [4], [6]. Lowering dopamine levels resulted in the *Neural* agent avoiding risks that lead to a higher payoff. These results are in agreement with a study in which a blockade of dopamine resulted in rats not making an extra effort of climbing over a barricade to get a high reward [17], and a study in which individuals with a polymorphism that lowers levels of dopamine in the prefrontal cortex tended to take less risks in a gambling task [7].

The model makes the following predictions: 1) The interaction between the DA and 5-HT neuromodulatory systems allows for appropriate decision making in games of conflict. In our model, when the VTA activity, which tracked expected reward, exceeded the Raphe activity, which tracked the expected cost, the agent would tend to escalate a fight (see Fig. 3). 2) Impairment to either the dopaminergic or serotonergic system will lead to perseverant, uncooperative behavior. In our model, impairment of the dopaminergic system resulted in risk-averse behavior (*Dove-like)* caused by an inability to assess reward, and impairment of the serotonergic system resulted in risk-taking behavior (*Hawk-like*) due to an inability to assess cost (see Table 3). 3) Although dopamine and serotonin activity appears to be related to different expectations (e.g., predictive reward, anticipated cost), the action of these neuromodulators on downstream targets is similar in that it governs decision-making. That is, phasic neuromodulation shifts an agent's behavior from random and exploratory to decisive and exploitive through differentially modulating synaptic pathways.

The model constructed for the present experiments is based on the notion that all neuromodulators have the same effect on downstream targets, but that specific neuromodulator levels are driven by environmental stimuli [10]. Large, phasic increases in neuromodulator activity cause an organism's behavior to be more exploitive or decisive, whereas lower levels of neuromodulatory activity result in the organism being more exploratory or indecisive. This is in agreement with the idea of cholinergic modulation of attention [18] and noradrenergic modulation of decision-making [19], but extends it to other neuromodulators such as dopamine and serotonin. Our model differs somewhat from the behavioral and neuroscience literature that suggests the role of dopamine is to calculate the reward prediction error, and that serotonin controls the timescale of the evaluation of delayed rewards in reinforcement learning [20], [21]. Instead it may be more in agreement with the proposal that neuromodulators, such as dopamine and serotonin are involved with the discovery of new actions to outcome mappings [22].

We designed our model to investigate how neuromodulation shapes behavior during competitive and cooperative situations. Our model has similarities to other computational models of neuromodulatory processes during decision-making [16], [20]; however, it tests a specific hypothesis of phasic neuromodulation, and applies it to game theory. Other computational models such as Evolutionary Algorithms and Reinforcement Learning have been effective in developing optimal strategies in games of

conflict [23], [24]. It may be of interest in the future to pit our neurobiologically inspired model against reinforcement learning and evolutionary algorithms.

Acknowledgements. This work was supported by the National Science Foundation (EMT/BSSE Award No.: 0829752) and the Office of Naval Research (Award No.: N000140910036).

References

1. Schultz, W., Dayan, P., Montague, P.R.: A neural substrate of prediction and reward. Science 275, 1593–1599 (1997)
2. Berridge, K.C.: Motivation concepts in behavioral neuroscience. Physiol. Behav. 81, 179–209 (2004)
3. Millan, M.J.: The neurobiology and control of anxious states. Prog. Neurobiol. 70, 83–244 (2003)
4. Crockett, M.J., Clark, L., Tabibnia, G., Lieberman, M.D., Robbins, T.W.: Serotonin modulates behavioral reactions to unfairness. Science 320, 1739 (2008)
5. Maynard Smith, J.: Evolution and the theory of games. Cambridge University Press, Cambridge (1982)
6. Wood, R.M., Rilling, J.K., Sanfey, A.G., Bhagwagar, Z., Rogers, R.D.: Effects of tryptophan depletion on the performance of an iterated Prisoner's Dilemma game in healthy adults. Neuropsychopharmacology 31, 1075–1084 (2006)
7. Roussos, P., Giakoumaki, S.G., Pavlakis, S., Bitsios, P.: Planning, decision-making and the COMT rs4818 polymorphism in healthy males. Neuropsychologia 46, 757–763 (2008)
8. Axelrod, R., Hamilton, W.D.: The evolution of cooperation. Science 211, 1390–1396 (1981)
9. Hasselmo, M.E., McGaughy, J.: High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. Prog. Brain Res. 145, 207–231 (2004)
10. Krichmar, J.L.: The Neuromodulatory System – A Framework for Survival and Adaptive Behavior in a Challenging World. Adaptive Behavior 16, 385–399 (2008)
11. Cox, B.R., Krichmar, J.L.: Neuromodulation as a Robot Controller: A Brain Inspired Design Strategy for Controlling Autonomous Robots. IEEE Robotics & Automation Magazine 16, 72–80 (2009)
12. Nowak, M., Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature 364, 56–58 (1993)
13. Amat, J., Paul, E., Zarza, C., Watkins, L.R., Maier, S.F.: Previous experience with behavioral control over stress blocks the behavioral and dorsal raphe nucleus activating effects of later uncontrollable stress: role of the ventral medial prefrontal cortex. J. Neurosci. 26, 13264–13272 (2006)
14. Watson, K.K., Ghodasra, J.H., Platt, M.L.: Serotonin transporter genotype modulates social reward and punishment in rhesus macaques. PLoS ONE e4156 (2009)
15. Cools, R., Roberts, A.C., Robbins, T.W.: Serotoninergic regulation of emotional and behavioural control processes. Trends Cogn. Sci. 12, 31–40 (2008)
16. Daw, N.D., Kakade, S., Dayan, P.: Opponent interactions between serotonin and dopamine. Neural Netw. 15, 603–616 (2002)

17. Denk, F., Walton, M.E., Jennings, K.A., Sharp, T., Rushworth, M.F., Bannerman, D.M.: Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. Psychopharmacology (Berl) 179, 587–596 (2005)
18. Pauli, W.M., O'Reilly, R.C.: Attentional control of associative learning-A possible role of the central cholinergic system. Brain Res. 1202, 43–53 (2008)
19. Aston-Jones, G., Cohen, J.D.: An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annual Review of Neuroscience 28, 403–450 (2005)
20. Doya, K.: Metalearning and neuromodulation. Neural Netw. 15, 495–506 (2002)
21. Schweighofer, N., Tanaka, S.C., Doya, K.: Serotonin and the evaluation of future rewards: theory, experiments, and possible neural mechanisms. Annals of the New York Academy of Sciences 1104, 289–300 (2007)
22. Redgrave, P., Gurney, K.: The short-latency dopamine signal: a role in discovering novel actions? Nature Reviews 7, 967–975 (2006)
23. Sandholm, T.W., Crites, R.H.: Multiagent reinforcement learning in the Iterated Prisoner's Dilemma. Biosystems 37, 147–166 (1996)
24. Harrald, P.G., Fogel, D.B.: Evolving continuous behaviors in the Iterated Prisoner's Dilemma. Biosystems 37, 135–145 (1996)