# Neuromodulation and Time-Dependent Plasticity in a Model of Foraging Behavior

Jeffrey L. Krichmar
Department of Cognitive Sciences
University of California, Irvine
Irvine, CA 92697 USA
jkrichma@uci.edu

*Abstract*— **In foraging behavior, where an animal searches for food caches, it is imperative for the animal to remember the locations and routes to these caches. An important consideration is the means by which the organism takes the appropriate actions to lead it to a goal that satisfies a particular need. We introduce a time-dependent plasticity rule that biases movement in a particular direction by developing asymmetric neuronal receptive fields through experience. The model contains hippocampal areas that respond differentially to locations in space, frontal cortex areas that respond to different salient cues from the environment, and neuromodulators that respond to rewards and costs. This model suggests a means by which neuromodulated time-dependent plasticity in the frontal cortex can facilitate action selection. It also suggests how these neuronal responses may lead to successful performance in a foraging task.**

***Frontal Cortex, Navigation, Place fields, Reinforcement, Search***

## I. INTRODUCTION

In foraging behavior, where pathways to food sources must be remembered, an animal selects a series of actions to acquire these sources [1]. The environment and the organism's internal state can influence these choices.

Theoretical and experimental work has identified the dopaminergic system as a possible neural correlate of reward expectation [2]. Alternatively, it has been suggested that the dopaminergic system signals the occurrence of novel events to facilitate action to outcome associations [3]. Based on these notions, reinforcement learning and temporal difference learning have been posited as a means to learn pathways to reward locations [4-6]. In these models, changes in expected reward reinforce actions that lead to goal locations.

An alternative means for assigning credit to actions that lead to goals may be brought about by experience-dependent changes to the shape of neuronal receptive fields. For example, hippocampal place fields may become asymmetric through LTP in the hippocampus in a way that leads to the prediction of motor sequences [7]. Mehta and colleagues showed that rat hippocampal place cells become skewed with experience and showed that this asymmetry could emerge through spike-timing dependent plasticity or STDP [8]. After experience, the shape of the place field became skewed such that activity increased as it approached a location and decreased sharply as it passed through that location. As a result, the rat's direction of travel could be inferred from the shape of these asymmetric place fields.

In this paper, we extend the notion of asymmetric hippocampal place fields to frontal cortical areas and replace the simple reward system used in previous modeling studies [4-7], with multiple reinforcers that encode different environmental cues and internal states. We present a model of neuromodulated time-dependent plasticity, in which a simulated agent learns pathways to rewards of varying saliency while avoiding obstacles. The type of reward it seeks depends on its internal state or drive. The model suggests a means by which frontal cortical areas and neuromodulators may work in concert to produce effective behavioral strategies.

## II. METHODS

### A. Simulation Environment and Experimental Protocol

The simulated agent explored a virtual environment that consisted of a 16x16 grid of locations (see Figure 1A). At the start of a trial, two rewards, two obstacles, and the agent were placed randomly at grid locations in the environment. To make the foraging task challenging, the two rewards were at least 8 locations away from each other. The two obstacles had a diameter of 4 grid locations. If the agent attempted to move into a location where an obstacle resided or beyond the boundaries of the environment, it stayed at its current location until the next simulation cycle. In each simulation cycle of the trial, the activities of the neurons, the strength of the connections, and the position of the agent were updated as described below. A trial lasted 2500 simulation cycles. At the start of the trial, the saliency for the first reward, S1, was set to 1 and the saliency for the second reward, S2, was set to 0. When the first reward, R1, was found, S1 was set to 0 and S2 was set to 1. The saliency for rewards toggled in this manner throughout the trial.

### B. Neural Simulation

A neural simulation was constructed that contained regions which loosely corresponded to the hippocampus and frontal cortex (see Figure 1B). The simulation contained a hippocampal area, *CA3*, which had place fields centered at locations in the virtual environment. Neurons in *CA3* projected to neurons in a *CA1* area that in turn projected to neurons in three different frontal areas; (1) *FCR1* – an area that responded to a particular reward type (*R1*) and salience (*S1*), (2) *FCR2* – an area that responded to reward *R2* and salience (*S2*), and (3) *FCD* – an area whose response was positively correlated with the agent's progress or distance covered. All of the synaptic

connections in the simulation were subjected to neuromodulated time-dependent plasticity (see Figure 2A). The neuromodulators could be reward related, $NM_R$, or related to the distance traveled by the agent, $NM_D$.

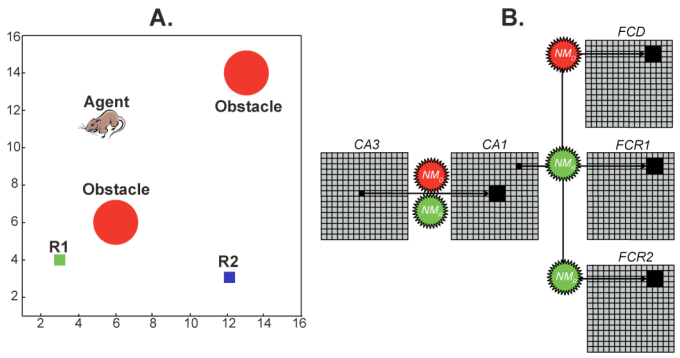


**Figure 1. A.** Foraging task environment consists of a 16x16 grid of locations, two types of rewards (R1 and R2), obstacles, and an impenetrable perimeter. **B.** Neural simulation architecture. The simulated nervous system consists of five neural areas each having 256 (16x16) neurons. There are two simulated hippocampal areas that respond to place; *CA3* and *CA1*, and there are three simulated frontal cortex areas that respond to saliencies, such as rewards; *FCR1* and *FCR2*; and distance traveled *FCD*. Each neuron in *CA3* responds preferentially, with a two-dimensional Gaussian tuning curve, to the location specified by its Cartesian coordinate (*x,y*). The synaptic connectivity of the network is uniform such that each pre-synaptic neuron, at a given Cartesian coordinate (*x,y*), projects to nine post-synaptic neurons centered around (*x,y*). All synaptic connections are subject to a neuromodulated time-dependent plasticity learning rule. The neuromodulators can be reward related ($NM_R$) or cost related based on distance traveled by the agent ($NM_D$).

Each neural area consisted of a two-dimensional grid of neurons (16x16), where each neuron corresponded to a location in the virtual environment. Each pre-synaptic neuron in one area projected to nine post-synaptic neurons in another area, centered on the corresponding location of the pre-synaptic neuron (see Figure 1B). All the neurons in the model had an activation function where activity ranged from 0 (i.e. quiescent) to 1 (i.e. maximally active).

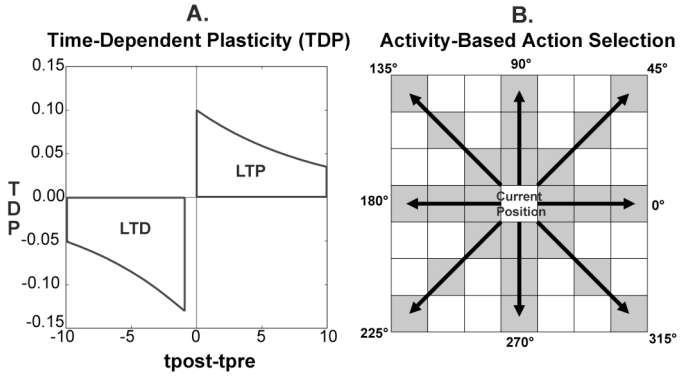


**Figure 2. A.** The timing dependent plasticity (*TDP*) function. The chart shows the amount of long-term potentiation (LTP) or long-term depression (LTD) that is applied to synaptic connection strength plasticity. The variables *tpost* and *tpre*, which represent simulation cycles, are indices into the times of pre- and post-synaptic activations of a neuron. The *TDP* rule ranges from 10 time steps forward (*tpost* ≥ *tpre*) to 10 steps backward (*tpre* < *tpost*). When *tpost* equals *tpre*, LTP is maximal at 0.10 and decays exponentially, with a time constant of -0.1, as *tpost-tpre* becomes larger. When tpost-tpre equals negative one, LTD is maximal at 0.13 and decays exponentially, with a time constant of -0.1, as *tpost-tpre* becomes smaller. **B.** Activity-based action selection. A new heading is chosen based on the summed activity of *FCD*, *FCR1*, and *FCR2*. The sum consists of the activity of three neurons, shown in grey, along one of eight headings from the current position. A softmax function is used to select between eight different headings. See text for details.

The *CA3* region of the model contained neurons that responded preferentially to locations in the environment. Each neuron had a unique location that it responded to with the maximum firing rate. The preferred locations of the 16x16 *CA3* neurons were such that they uniformly covered the environment. The firing rate ranged from maximally active at the preferred location to quiescent at locations far from the preferred one according to the following Gaussian tuning curve function:

$$CA3(i,t+1) = \exp\left(\frac{-\left(\left\|p(t)-s_i\right\|\right)^2}{2}\right) + rnd(-0.05,+0.05) \quad (1)$$

where *i* specifies the neuron in the *CA3* area, *t* is the simulation cycle, *p(t)* represents the location of the agent in the environment, $s_i$ represents the preferred location of *CA3* neuron *i*, and *rnd* returns a random number uniformly distributed from -0.05 to 0.05.

*CA1* neural activity depended on synaptic input from the *CA3* region. A *CA1* neuron's activity was calculated based on the following equation:

$$CA1(i,t+1) = \tanh\left(\sum_{j=1}^{9} CA3(j,t) * w_{ca3\to ca1}(i,j,t)\right) + rnd(-0.05,+0.05) \quad (2)$$

where *i* specifies the neuron in the *CA1* neural area, *j* is the index into the *CA3* neural area, *w* is the synaptic weight from *CA3* neuron *j* to *CA1* neuron *i*.

*FCR1* neural activity depended on the agent's internal state (*S1*), its proximity to a reward (*R1*), and synaptic input from the *CA1* region. In the agent simulations, it was assumed that the agent could sense a nearby reward and this would lead to activation of *FCR1*. Each *FCR1* neuron's activity was calculated by the following equation:

$$FCR1(i,t+1) = \tanh\left(S1 * \exp\left(\frac{-\left(\left\|p(t)-s_i\right\|\right)^2}{8}\right) * \exp\left(\frac{-\left(\left\|p(t)-R1\right\|\right)^2}{8}\right) + \left(\sum_{j=1}^{9} CA1(j,t) * w_{CA1\to FCR1}(i,j,t)\right)\right) + rnd(-0.05,+0.05) \quad (3)$$

where *S1* is the saliency drive for reward *R1*, and *w* is the synaptic weight from *CA1* neuron *j* to *FCRC1* neuron *i*.

*FCR2* neural activity was determined similar to a *FCR1* neuron except that the response is based on saliency *S2* and reward *R2*.

*FCD* neural activity depended on synaptic input from the *CA1* region. The activity function of a *FCD* neuron was calculated by the following equation:

$$FCD(i,t+1) = \tanh\left(\sum_{j=1}^{9} CA1(j,t) * w_{CA1\to FCD}(i,j,t)\right) + rnd(-0.05,+0.05) \quad (4)$$

where *w* is the synaptic weight between *CA1* neuron *j* to *FCD* neuron *i*.

### C. Synaptic Plasticity

All synaptic connections in the model were subject to an experience-dependent plasticity rule that took into account the timing of pre- and post-synaptic activity (see Figure 2A), as well as amplification of learning through simulated neuromodulators (see Figure 1B). In this learning rule, the direction of change in connection strength between a pre- and post-synaptic neuron was determined by the timing of activity similar to STDP [9, 10]. However, in the time-dependent plasticity rule presented here the magnitude of change was based on firing rate as opposed to spikes. Furthermore, the level of simulated neuromodulators dictated synaptic change.

Neuromodulator levels were determined by a reward evaluation function ($NM_R$) and a distance traveled cost function ($NM_D$).

$NM_R$ was set either due to the agent being at the reward or due to the activity of neural areas *FCR1* and *FCR2*.

$$NM_R(t+1) = \begin{cases} 3; & if \left\| R1 - p(t) \right\| < 1 \, and \, S1 = 1 \\ 3; & if \left\| R2 - p(t) \right\| < 1 \, and \, S2 = 1 \\ \left( FCR1(i,t) + FCR2(i,t) \right) - 0.10; & otherwise \end{cases} \quad (5)$$

where $p(t)$ represents the location of the agent at time $t$, $i$ represents the index into the neural areas that corresponds to location $p$.

The $NM_D$ was proportional to the distance traveled by the agent.

$$NM_D(t+1) = \frac{\left\| p(t) - p(t - \Delta t) \right\|}{\Delta t} - 0.10 \quad (6)$$

where $\Delta t$ is equal to 10 simulation cycles.

The weights from *CA3* to *CA1* were modulated by both $NM_R$ and $NM_D$, the weights from *CA1* to *FCR1* and *FCR2* were modulated by $NM_R$, and the weights from *CA1* to *FCD* were modulated by $NM_D$ (see Figure 1B). The change in weight was calculated as follows:

$$w(i,j,t+1) = w(i,j,t) - 0.0001 + \delta * NM * \sum_{tpre=t-T}^{t} \sum_{tpost=t-T}^{t} a(j,tpre) * a(i,tpost) * TDP(tpost - tpre) \quad (7)$$

where $t$ is the index into the current simulation cycle, $i$ is the index of the post-synaptic neuron, $j$ is the index of the pre-synaptic neuron, $\delta$ is the learning rate which is set to 0.25 for *learning* agents and 0.0 for the *no learning* agents, *tpre* and *tpost* are simulation cycle indices, $T$ is equal to 10, $a(j,tpre)$ is the activity level of pre-synaptic neuron $j$, $a(i,tpost)$ is the activity of the post-synaptic neuron $i$, *TDP* is the time-dependent plasticity function that dictates the direction and magnitude of change (see Figure 2A), and *NM* is the level of neuromodulator in the system.

### D. Action Selection

After each simulation cycle, the agent selected a new heading and position based on the summed activity of the *FCD*, *FCR1*, and *FCR2* neural areas. From this combined activity, eight population vectors were calculated by summing over the activity of three neurons radiating from the current position toward a one of eight headings (see Figure 2B). A softmax algorithm was used to create a probability distribution of possible headings from these population vectors:

$$p_{hdg} = \frac{\exp(2a_{hdg})}{\sum_{h=1}^{8} \exp(2a_h)}; \quad (8)$$

where *hdg* and *h* represent one of 8 headings, $p_{hdg}$ is the probability of selecting a new directional heading, is $a_{hdg}$ and $a_h$ are the summed activities of *FCD*, *FCR1*, and *FCR2* for a given heading (see Figure 2B). A new heading was selected based on the probabilities derived from this softmax function. Similar to Foster and colleagues [5], momentum of the agent's heading was modeled by a mixture of 25% from the softmax function and 75% of the previous heading. This restricted the turning curve of the agent, and was particularly important early on, when the whole environment must be searched fairly quickly. The new $(x,y)$ position of the agent was calculated from the cosine and sine of the new heading.

## III. RESULTS

The simulated agent's task was to find two different reward sources, avoid obstacles, and circumvent impasses (see Figure 1A). It was assumed that the agent needed only one reward type at a time and that after finding that reward source it was sated. This need was modeled by the agent's internal state. For example, initially the agent's internal state dictated that it was "hungry" and in need of food. After finding the food source, the agent's internal state dictated that it was "thirsty" and needed water.

The simulated agent's ability to learn an effective search strategy in a foraging task was judged by comparing the simulated agent's performance with learning to the performance of a group of agents that had learning disabled. Experiments with both the *learning* group and the *no learning* group consisted of 100 trials of 2500 simulation cycles each.

### A. Behavioral Results

The *learning* group found significantly more rewards than the group with learning disabled (see Figure 3). Lesions of either the *FCR* or *FCD* frontal areas significantly degraded the agent's performance. The superior performance of the *learning* group was reflected in its search strategy. Search metrics, which are used to analyze foraging performance of biological organisms [11-13], were applied to the performance of the model. The *learning* group's trajectory was significantly straighter than the *no learning* group (see Straightness in Figure 4), and its foraging was more focused than the *no learning* group (see Thoroughness in Figure 4). Figure 5 shows trajectories from two different trials in which the agent shuttles between different reward sources while avoiding obstacles.

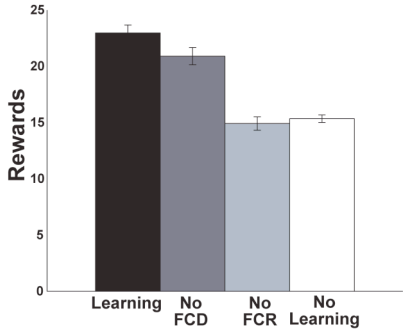


**Figure 3.** Average number of rewards per trial under different conditions. Error bars denote standard error. The number of rewards found by the *Learning* group were significantly larger than the *No Learning* group ($p << 0.001$, Wilcoxon rank sum test), than the *No FCD* group ($p < 0.02$, Wilcoxon rank sum test), and than the *No FCR* group ($p << 0.001$, Wilcoxon rank sum test).

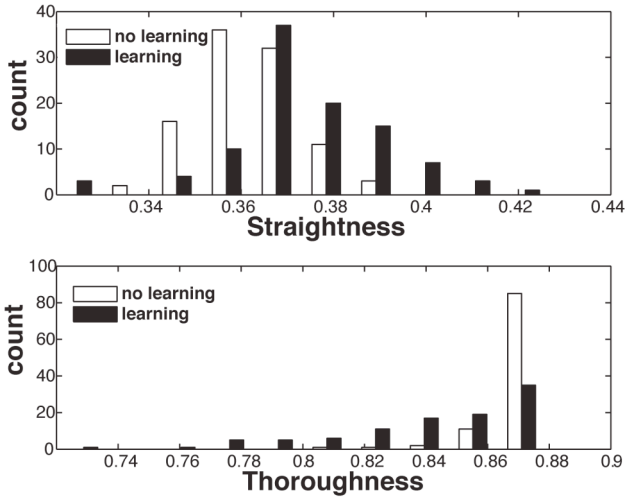


**Figure 4.** Search metrics for the performance of the simulated agents. The path trajectory was significantly straighter for the *Learning* group than the *No Learning* group ($p << 0.001$, two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis test). The thoroughness of the exploration was significantly smaller for the *Learning* group than the *No Learning* group ($p << 0.001$, two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis test).

### *B. Neuronal Responses During Behavior*

The trajectories of the agents, shown in Figure 5, were reflected in the patterns of connection strengths that developed due to neuromodulated time-dependent plasticity (see Figure 6). Pathways that led to rewards were reflected by strong biases in connection strengths toward those rewards (see *CA1→FCR1* and *CA1→FCR2* in Figure 6). Moreover, pathways in which the agent made progress or covered distances effectively were reflected in strong biases in connection strengths that led to increased neural activity in locations away from obstacles and borders (see *CA1→FCD* in Figure 6). The route the agent took could be inferred by combining these weights (see Sum of Weights in Figure 6).

The weight biases were brought about by the neuromodulated *TDP* learning rule (see Methods, and Figure 2A) and resulted in asymmetric place fields. A measure of asymmetry, the Firing Rate Asymmetry Index (*FRAI*), was calculated for the neuron's receptive fields:

$$FRAI = \frac{F1 - F2}{F1 + F2}; \quad (9)$$

where *F1* is the sum of activity of a neuron during the three time steps prior to reaching the neuron's initial preferred location, and *F2* is the sum of activity of a neuron during the three time steps after passing through the neuron's initial preferred location. A value of *FRAI* close to zero would signify a symmetric receptive field, whereas a positive value of *FRAI* would denote anticipatory activity as the agent approaches a location.

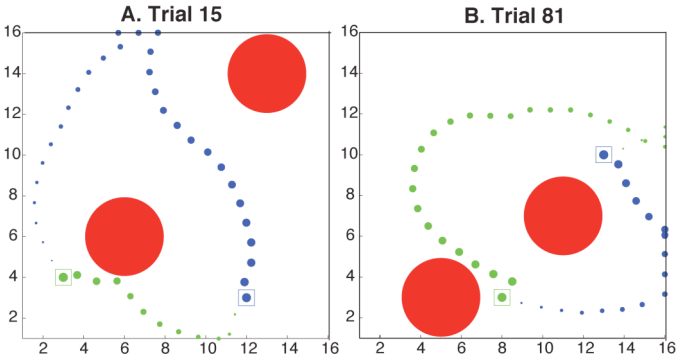


**Figure 5.** Sample trajectories of two representative agents. The first reward, corresponding to one salience, is denoted by an open green square, and the second reward, corresponding to another salience, is denoted by an open blue square. Obstacles are shown as red circles. The filled circles denote the agent's position and the diameter of the circles denote the agent's trajectory from beginning (smaller circles) to finding the reward (largest circles).

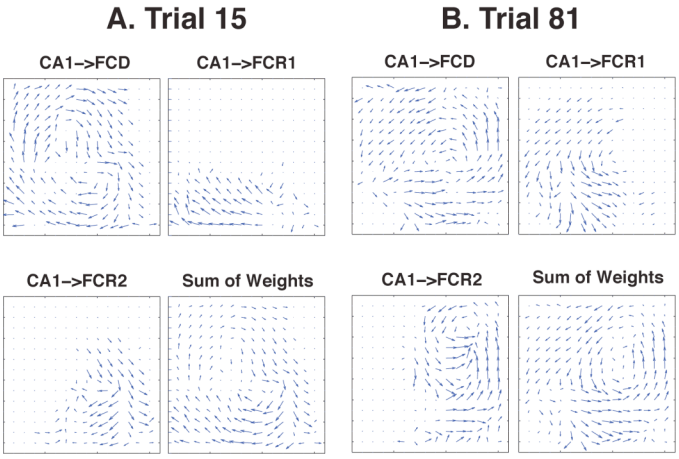


**Figure 6.** Connection strengths for the synaptic projections in the neural simulation. The arrows in the figures denote the vector summation of the synaptic weights from a pre-synaptic neuron (tail of arrow) to the nine post-synaptic neurons. The direction of the arrow points to the most strongly connected post-synaptic neurons and tends to predict the direction of simulated agent's movement. **A.** Connection strength vectors for a representative simulation trial (Trial 15). **Upper-left.** *CA1* to *FCD*. **Upper-right.** Connection strengths from *CA1* to *FCR1*. **Lower-left.** Connection strengths from *CA1* to *FCR2*. **Lower-right.** Summation of the weight vectors *CA1→FCD, CA1→FCR1,* and *CA1→FCR2*. **B.** Connection strength vectors for a representative simulation trial (Trial 81).

Experience resulted in a skewing of receptive fields such that neuronal activity was anticipatory of a future location (see Figure 7). The place fields of neurons in *CA3*, which did not have plastic connections, stayed symmetric throughout the trial. However, neurons that did have plastic connections (e.g. *CA1*,

*FCD*, *FCR1*, and *FCR2*) became significantly positively skewed with experience. This predictive response arises from the learning rule (see Equation 7), in which weights were potentiated when the pre-synaptic activity preceded the post-synaptic activity on pathways that led the agent toward rewards or unimpeded progress.

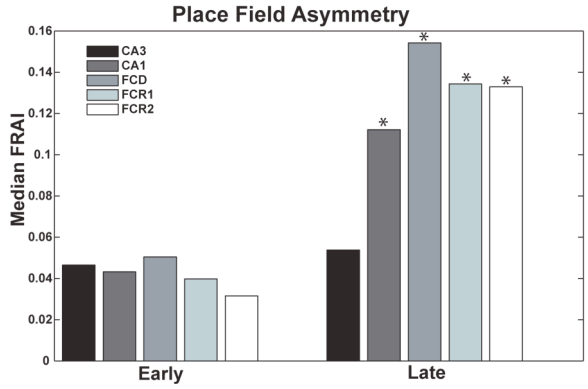


**Figure 7.** Median Firing Rate Asymmetry Indices (*FRAI*; see equation 9) for neural areas in the model. The *FRAI*s are compared between the first quarter of a trial (*Early*) and the last quarter of a trial (*Late*). With the exception of *CA3* ($p > 0.01$; Wilcoxon rank sum), the *Late* FRAIs were significantly larger, denoted by an asterisk, than the *Early* FRAIs ($p << 0.01$; Wilcoxon rank sum).

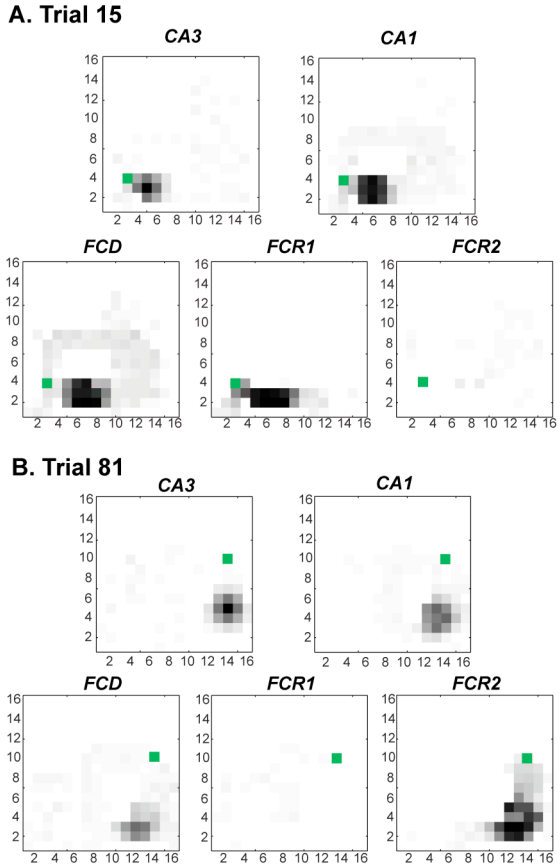


**Figure 8.** Rate maps for neurons in different neural areas. Each pixel indicates the mean activation of the neuron when the agent was in that particular location of its environment. The activation level is denoted by a grey scale where white represents no activity and black represents maximum activity. **A.** Rate maps for neurons that initially have a preferred location at coordinates (5,3) in the neural areas *CA3*, *CA1*, *FCD*, *FCR1*, and *FCR2* during Trial 15. The green square denotes the location of the reward (R1) the agent is approaching. **B.** Rate maps for neurons that initially have a preferred location at coordinates (14,5) in the neural areas *CA3*, *CA1*, *FCD*, *FCR1*, and *FCR2* during Trial 81. The green square denotes the location of the reward (R2) the agent is approaching. Note that the neuron in *CA3* is symmetric about its location in space, whereas other areas have developed asymmetric place fields.

Figure 8 shows representative examples of asymmetric place fields. In Figure 8A, the neurons have initial preferred locations at coordinates (5,3). After experience these receptive fields, with the exception of *CA3*, became predictive of where the agent intended to move; in this case the reward location marked in green. Note that since the agent was searching for reward *R1*, the neural area *FCR2* is relatively inactive. Similarly, receptive fields in Figure 8B show a shift from their initial preferred location of (14,5) that is anticipatory of the *R2* reward location marked in green.

After experience in the environment, the path selected by the agent, which was based on the population activity of the *FC* neural areas, became biased towards rewards (see Figure 9). After foraging experience, the center of population activity tended to be away from the agent's current position (red marker in Figure 9) and towards a reward location (green marker in Figure 9). The shift of activity towards the reward is a direct consequence of the changes in synaptic plasticity with stronger weights pointing towards reward locations (see Figure 6) and the asymmetry of the place fields was such that the neurons fire in anticipation of moving towards a reward location (see Figure 8).

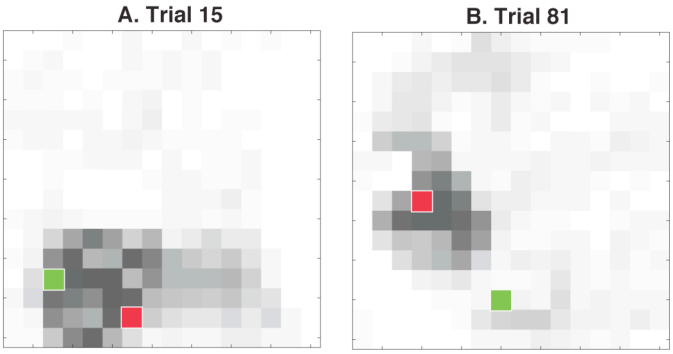


**Figure 9.** Summation of neural activity of areas *FCD*, *FCR1*, and *FCR2* when simulated agent was in a given location (denoted by the red pixel). Each pixel indicates the summated activation of a neuron, where the activation level is denoted by a grey scale where white represents no activity and black represents maximum activity. The green pixel denotes the location of a reward. Note that the population of activity tends to be skewed toward the reward location. **A.** Summation of activity towards the end of Trial 15. **B.** Summation of activity towards the end of Trial 81.

Taken together, these results show that a learning rule that took into account the temporal dynamics of neural activity, combined with modulators that were tied to particular saliencies of the environment and the agent's internal state were sufficient for learning routes between reward locations and around obstacles.

## IV. DISCUSSION

We have shown that a model having a learning rule dependent on the timing of neuronal activity and the

neuromodulation by multiple reinforcers, can develop effective search strategies for foraging behavior (see Figures 3 through 5). In this model, there was reinforcement from two different reward types, as well as a cost function in which reinforcement was proportional to distance traveled. The agent discovered reward locations and passable areas of the environment through experience-dependent learning. Together, these reinforcement signals were sufficient for shaping the neuronal connectivity such that the agent navigated towards a particular goal depending on its internal state while avoiding obstacles (see Figure 6). Navigation performance was evaluated by not only the rewards found, but also by the manner in which the agent explored its environment (see Figure 4). The *learning* agent took straighter paths (i.e. higher straightness values) and had a more focused search pattern (i.e. lower thoroughness values) than the *no learning* agents and the agents with lesions to its frontal cortex.

In our model, the routes leading to rewards did not rely on changes in expected value [2, 5], but rather on directionally asymmetric neural activity brought about by time-dependent plasticity and exploration [7]. There is evidence, from rodent hippocampal recordings, that experience can alter the shape of neuronal activity from symmetric about a place to skewed in the direction of a familiar pathway [8]. Our model incorporated a time-dependent plasticity rule, which was sufficient to generate asymmetric place fields hippocampal region of the model, and extended this notion to the frontal cortical regions of the model.

The asymmetric neuronal responses brought about by time-dependent plasticity suggest that activity in the frontal cortex may be skewed such that the response becomes predictive. There is evidence of anticipatory activity in the frontal cortex that leads to decisions in simple discrete choice tasks [14-16]. Furthermore, there is evidence that different areas of the frontal cortex encode different reward types, as well as the cost of a decision [17-20].

The model also assumes that different neuromodulatory systems respond to different environmental stimuli. Besides the dopaminergic (DA) system, which appears to drive reward anticipation [2], other neuromodulators may be related to the cost neuromodulator ($NM_D$) in the present model. For example, the serotonergic (5-HT) system influences risk aversion [21], the cholinergic (ACh) system is related to the level of attentional effort [22], and the noradrenergic (NE) system appears to respond to novel or salient objects [23].

The model presented here suggests how different areas of the forebrain, which evaluate different rewards and costs, may work in concert with neuromodulatory systems to develop effective behavioral repertoires.

## ACKNOWLEDGMENT

The author would like to thank G. Edelman, J. Gally and A. Seth for useful suggestions and discussions. The author would also like to thank R. Martin for assistance in software development. This work was supported by grants from the Office of Naval Research, Defense Advanced Research Projects Agency, and the Neurosciences Research Foundation.

## REFERENCES

[1] T. J. Prescott, J. J. Bryson, and A. K. Seth, "Introduction. Modelling natural action selection," *Philos Trans R Soc Lond B Biol Sci,* Apr 11 2007.

[2] W. Schultz, P. Dayan, and P. R. Montague, "A neural substrate of prediction and reward," *Science,* vol. 275, pp. 1593-9, Mar 14 1997.

[3] P. Redgrave and K. Gurney, "The short-latency dopamine signal: a role in discovering novel actions?," *Nat Rev Neurosci,* vol. 7, pp. 967-75, Dec 2006.

[4] A. Arleo, F. Smeraldi, and W. Gerstner, "Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning," *IEEE Trans Neural Netw,* vol. 15, pp. 639-52, May 2004.

[5] D. J. Foster, R. G. Morris, and P. Dayan, "A model of hippocampally dependent navigation, using the temporal difference learning rule," *Hippocampus,* vol. 10, pp. 1-16, 2000.

[6] J. L. Krichmar, A. K. Seth, D. A. Nitz, J. G. Fleischer, and G. M. Edelman, "Spatial navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions," *Neuroinformatics,* vol. 3, pp. 197-221, 2005.

[7] K. I. Blum and L. F. Abbott, "A model of spatial map formation in the hippocampus of the rat," *Neural Comput,* vol. 8, pp. 85-93, Jan 1996.

[8] M. R. Mehta, M. C. Quirk, and M. A. Wilson, "Experience-dependent asymmetric shape of hippocampal receptive fields," *Neuron,* vol. 25, pp. 707-15, Mar 2000.

[9] W. Senn, H. Markram, and M. Tsodyks, "An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing," *Neural Comput,* vol. 13, pp. 35-67, Jan 2001.

[10] S. Song, K. D. Miller, and L. F. Abbott, "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity," *Nat Neurosci,* vol. 3, pp. 919-26, Sep 2000.

[11] A. Bond, "Optimal foraging in a uniform habitat: the search mechanism of the green lacewing," *Animal Behavior,* vol. 28, pp. 10-19, 1980.

[12] P. M. Kareiva and N. Shigesada, "Analyzing Insect Movement as a Correlated Random Walk," *Oecologia,* vol. 56, pp. 234-238, 1983.

[13] D. L. Winkelman and G. L. Vinyard, "Gyrinid searching tactics: empirical observations and a tactical model," *Behavioral Ecology Sociobiology,* vol. 28, pp. 345-251, 1991.

[14] J. I. Gold and M. N. Shadlen, "Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward," *Neuron,* vol. 36, pp. 299-308, Oct 10 2002.

[15] J. D. Schall, "Neural basis of deciding, choosing and acting," *Nat Rev Neurosci,* vol. 2, pp. 33-42, Jan 2001.

[16] P. L. Smith and R. Ratcliff, "Psychology and neurobiology of simple decisions," *Trends Neurosci,* vol. 27, pp. 161-8, Mar 2004.

[17] M. M. Botvinick, J. D. Cohen, and C. S. Carter, "Conflict monitoring and anterior cingulate cortex: an update," *Trends Cogn Sci,* vol. 8, pp. 539-46, Dec 2004.

[18] J. O'Doherty, M. L. Kringelbach, E. T. Rolls, J. Hornak, and C. Andrews, "Abstract reward and punishment representations in the human orbitofrontal cortex," *Nat Neurosci,* vol. 4, pp. 95-102., 2001.

[19] P. H. Rudebeck, M. E. Walton, A. N. Smyth, D. M. Bannerman, and M. F. Rushworth, "Separate neural pathways process different decision costs," *Nat Neurosci,* vol. 9, pp. 1161-8, Sep 2006.

[20] M. F. Rushworth, T. E. Behrens, P. H. Rudebeck, and M. E. Walton, "Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour," *Trends Cogn Sci,* vol. 11, pp. 168-76, Apr 2007.

[21] M. J. Millan, "The neurobiology and control of anxious states," *Prog Neurobiol,* vol. 70, pp. 83-244, Jun 2003.

[22] M. G. Baxter and A. A. Chiba, "Cognitive functions of the basal forebrain," *Curr Opin Neurobiol,* vol. 9, pp. 178-83, Apr 1999.

[23] A. J. Yu and P. Dayan, "Uncertainty, neuromodulation, and attention," *Neuron,* vol. 46, pp. 681-92, May 19 2005.