Affective Cognitive Learning and Decision Making: The Role of Emotions

Hyungil Ahn and Rosalind W. Picard
MIT Media Laboratory
Cambridge, MA 02139, USA
{hiahn, picard}@media.mit.edu

Abstract

This paper presents a new computational framework where both 'the extrinsic reward from the external goal or cost' and 'the intrinsic reward from multiple emotion circuits and drives' play an integral role in learning and decision making. We show that the integration of the intrinsic reward from affect systems can be used for enhancing the efficacy of learning and decision making. In particular, we suggest a model of the affective anticipatory reward that is assumed to arise from the emotional seeking system. Our simulation results for a single-step choice and sequential multi-step choices show that affective biases from affective anticipatory rewards can be applied for improving the speed of learning, regulating the trade-off between exploration and exploitation in learning more efficiently, and adjusting the weight given to the immediate rewards over the future rewards in obtaining a decision making policy.

1 Introduction

Recent research in psychology, neuroscience, cognitive science, and neuroeconomics has shown that emotions play a key role in learning and decision making.

In particular, affective neuroscience has shown that the valenced affective feeling states provide fundamental values for the guidance of behavior, and the seeking system is significant in the emotional function of the brain - the basic impulse to search, investigate, and make sense of the environment [7]. Also, it has been reported that the covert biases related to previous emotional experience of comparable situations assist the reasoning process and facilitate the efficient processing of knowledge and logic necessary for conscious decisions [1].

Researchers in affective neuroscience have argued that rewards come from internal feelings, and these can arise from intrinsic causes as well as extrinsic ones. Also, work in neuroeconomics has incorporated emotions in models of decision making to explain human decision making under risk and intertemporal choice: for example, Loewenstein and O'Donoghue [4] have suggested a decision-making process in which a person's behavior is the outcome of interaction between a deliberative system that assesses options with a broad, goal-based perspectives and an affective system that encompasses emotions and motivational drives.

As far as we know, most work in artificial intelligence and machine learning has considered only the extrinsic motivational reward that arises from the external goal or cost. However, there have been some efforts to integrate a cognition model with a kind of intrinsic motivation model: Oudeyer and Kaplan [6] have modeled intelligent adaptive curiosity as a source of self-development. Singh et al. [8] have proposed a model of intrinsically motivated reinforcement learning based on the options framework. Also, Morén and Balkenius [5] have suggested a framework that ties together emotions, motivations and action selection, based on Mowrer's two process theory of learning.

This paper presents a new computational framework where both extrinsic and intrinsic rewards play an integral role in learning and decision making. We also suggest a model of affective anticipatory reward that is assumed to arise from the emotional seeking system. Our simulation results for a single-step choice such as a stochastic two-armed bandit type gambling task and sequential multi-step choices such as a maze task show that affective biases from affective anticipatory rewards can be applied for enhancing the efficacy of learning and decision making.

2 Types of Motivational Rewards

A lot of researchers in artificial intelligence and machine learning have been interested in the problem of 'how the motivational rewards influence learning and decision making', rather than the problem of 'where the motivational rewards arise from'. However, we think that it is important to know where the rewards arise from and how we can decompose those rewards and model each of them.

Neuroscientists Berridge and Robinson [2] have decomposed motivational reward into implicit incentive salience 'wanting' and explicit cognitive incentive expectations. They proposed the concept of incentive salience ('wanting') based on the findings that manipulation of dopamine systems powerfully changes

motivated behavior. The incentive salience 'wanting' refers to an underlying implicit and objective motivation process. They also describe explicit cognitive incentives: known or imagined (cognitive incentive representation), expected to be pleasant (hedonic expectation), subjectively desired and intended to be gained, or known to be obtainable by actions that cause it to occur (understanding of act-outcome causality). Such cognitive incentives are relevant for goal-directed strategies of action.

Loewenstein and Lerner [3] have proposed two different ways in which emotions influence decision making: expected emotions and immediate emotions. Expected emotions consist of predictions about the emotional consequences of decision outcomes. Dominant models of decision making, such as the expected utility model, assume that people attempt to predict the emotional consequences associated with alternative courses of action and then select actions that maximize positive emotions and minimize negative emotions. The second kind of affective influence on decision making consists of immediate emotions that are experienced at the time of decision making. Immediate emotions reflect the combined effects of anticipatory influences and incidental influences: the former arises from contemplating the consequences of the decision itself and the latter arises from factors unrelated the decision.

In the proposed computational framework we decompose rewards into 'the extrinsic reward from the external goal or cost' and 'the intrinsic reward from multiple emotion circuits and drives'. We take the position that emotion is a core critic in reward learning for evaluating environmental stimuli to cope with a variety of survival needs. Moreover, there is not just one circuit for providing emotional value, but Panksepp and others have demonstrated the existence of multiple value-generating systems in mammalian brains [7]. For example, Panksepp emphasizes basic emotion circuits such as a seeking circuit, a play circuit, a rage circuit, a fear circuit, a distress circuit, and others.

We have begun to build a computational model that tries to represent the above ideas by combining a set of intrinsic emotion-like circuits with more traditionally derived extrinsic value representations. While the emotion-like circuits should probably be multiple (as in Panksepp) we begin here by focusing on one of the circuits, the seeking circuit, which is also very similar to Berridge's 'wanting' circuit. The seeking system contributes to our feelings of engagement and excitement as we seek the resource or goal needed for our emotional and bodily states, and also when we pursue the cognitive interests that bring positive existential meanings into our lives [7]. This circuit is also called the expectancy system, the behavioral activation system, or the incentive or appetitive motivational system that mediates 'wanting'.

3 The Algorithm

First, we want to define an affective state and a cognitive state for the use in the algorithm. Suppose that a person makes a decision $d \in D$ for a given vector of environmental stimuli. These stimuli can activate the valenced affective feeling states through multiple emotion circuits and drive systems. We assume that an affective state a denotes these multiple affective feeling states.

Each affect system ($\theta \in A = \{0, 1, \cdots, |A|\}$), such as emotion circuits or drive systems, is represented by a pair of probabilities $a_\theta = (a_{\theta,1}, a_{\theta,2})$: $a_{\theta,1}$ and $a_{\theta,2}$, respectively, denote the conditional probabilities of feeling bad and feeling good. The detail is described below.

For simplicity in the current model, we will assume that the sum of the two probabilities is always one. In this paper an affective state is represented by a vector of multiple pairs of probabilities $a = \{a_\theta\} = (a_0, a_1, \cdots, a_{|A|})$. In particular, we focus on the affective anticipatory reward from the seeking circuit ($\theta = 0$) in this paper. The two probabilities of feeling bad or good in the seeking circuit are denoted by $a_0 = (a_{0,1}, a_{0,2})$.

We assume that a cognitive state also activated by a vector of environmental stimuli denotes a vector representation of 'a belief state of the world related to the goal of decision making' or 'memories of broader goals [4]'. However, since we only deal with the fully observable world in this paper, a cognitive state c can be assumed to be a fully observed state of the world.

To explain the cognition, affect and decision-making models, let c, a, d, c' and a', respectively, denote dummy variables for a current cognitive state (c), a current affective state (a), a current decision (d) being chosen, a next cognitive state (c') and a next affective state (a') after acting upon the current decision.

The cognition model $\Pr(c' \mid c, d)$ is a cognitive state transition model. In this paper this model can be purely learned by experience and considered a simulation model, but later we will be able to try to incorporate the prior knowledge or other models of cognitive representations into this model. We think that the role of cognition in learning and decision making is very related to complex information processing, future predictions for consequences of a current action and consideration of future delayed rewards in decision making.

The affect model for each emotion circuit or drive system is $a'_\theta = \Pr(v'_\theta \mid c', c, a_\theta, d)$ where v'_θ can take one of two possible states: the feeling-bad state ($v'_\theta = 1$) or the feeling-good state ($v'_\theta = 2$). In order to model each affect system, it is important to know more about the unconditioned stimuli and how classical conditioning works to associate the neutral or conditioned stimuli with the unconditioned stimuli for evaluating a biological value. In this paper we only focus on the model of the seeking circuit and also, we assume that the affective anticipatory reward that arises from the emotional seeking circuit can be represented by the two components of valence and arousal, and purely learned by experience. Thus, valence due to seeking is captured by $a = a_0 = (a_{0,1}, a_{0,2}) = (\Pr(\text{feeling bad while seeking}), 1 - \Pr(\text{feeling bad while seeking}))$. More details of how

this is combined with arousal are given below.

The decision-making model $\Pr(d \mid c, a)$ computes the probability of choosing a decision d conditioned on the current cognitive state c and the current affective state a, and uses the Boltzmann selection that can easily control the trade-off between exploration and exploitation through the inverse temperature β.

The following algorithm describes an on-line affective and cognitive learning and decision-making algorithm that runs the loop from (STEP 1) to (STEP 8) repeatedly[1].

In (STEP 1), $R_0(v'_0 = 1)$ is -1 (the unit reward for the feeling-bad state (1)) and $R_0(v'_0 = 2)$ is $+1$ (the unit reward for the feeling-good state (2)). In addition, η_0 is a varying scale factor controlling the relative influence of the immediate affective anticipatory reward and the long-run rewards. As η_0 increases, the influence of the immediate affective anticipatory reward gets strong. Loewenstein and O'Donoghue [4] have mentioned that this kind of scale factor is closely related to the cost of willpower.

We model the affective anticipatory reward by $\sigma_0(c^t, d^t) \sum_{c'=1}^{|C|} \sum_{v'_0=1}^{2} R_0(v'_0)\ \Pr(v'_0|c', c^t, a^t_0, d^t) \Pr(c'|c^t, d^t)$ in (STEP 1) by valence and arousal for the current choice. The arousal increases with increasing uncertainty and is modeled by $\sigma_0(c^t, d^t)$. The valence such as feeling good or bad is modeled by the rest part which is positive when the current choice is expected to give the reward greater than the average $R_{mean}(c^t)$, and negative when the choice is expected to give the reward less than the average. Depending on whether valence is good or bad, the uncertainty can be interpreted as hope (feeling lucky) or uneasiness (feeling risky).

In (STEP 4), when δ_E is positive, the affective state for the same condition (the decision d^t and the cognitive state c^{t+1}) next time has a higher probability of feeling good. Otherwise, the affective state for the same condition next time has a higher probability of feeling bad.

In (STEP 5), $\sigma_0(c, d)$ computes the standard deviation of the extrinsic reward distribution when a current cognitive state is c and a current decision is d. We assume that this value represents the uncertainty level of choosing the decision at the current cognitive state.

.....

(STEP 1) For the current cognitive state $c^t \in C = \{1, \cdots, |C|\}$ and the current affective state a^t, choose a decision $d^t \in D = \{1, \cdots, |D|\}$ from the following distribution and act upon the decision d^t:

$$\Pr(d^t|c^t, a^t) := \frac{\exp(\beta Q_{DM}(c^t, a^t, d^t))}{\sum_{d=1}^{|D|} \exp(\beta Q_{DM}(c^t, a^t, d))}$$

[1](i) In particular, we focus on the affective anticipatory reward from the seeking circuit ($\theta = 0$) in this paper.
(ii) c, a, d, c' and a' are all dummy variables.

For all $d \in D$, the decision value Q_{DM} (= the extrinsic value Q_{ext} + the intrinsic value from affect systems $\theta \in A = \{0 \text{ (seeking)}, 1, \cdots, |A|\}$)

$$Q_{DM}(c^t, a^t, d) := Q_{ext}(c^t, d) + \sum_{\theta=0}^{|A|} \eta_\theta\ \sigma_\theta(c^t, d) \sum_{c'=1}^{|C|} \sum_{v'_\theta=1}^{2} R_\theta(v'_\theta)\ \Pr(v'_\theta|c', c^t, a^t_\theta, d) \Pr(c'|c^t, d)$$

(STEP 2) When the decision d^t is acted upon, activate a new cognitive state $c^{t+1} \in C = \{1, \cdots, |C|\}$ for new environmental stimuli and obtain the extrinsic reward r_{ext}

(STEP 3) The cognition model $\Pr(c'|c, d)$, the extrinsic reward models $R_{ext}(c'|c, d)$, $S_{ext}(c'|c, d)$, and the average reward model $R_{mean}(c)$ are updated by using an experience tuple $< c^t, d^t, r_{ext}, c^{t+1} >$. $R_{mean}(c)$ is the average reward for probably the two best choices (d_1 and d_2) with which the greatest or the second greatest mean reward was given. $f(c, d, c')$ denotes the number of visits to c' from c after d was acted upon:

$$f(c^t, d^t, c^{t+1}) \leftarrow f(c^t, d^t, c^{t+1}) + 1$$

$$\Pr(c'|c^t, d^t) := \frac{f(c^t, d^t, c')}{\sum_{k=1}^{|C|} f(c^t, d^t, k)} \quad \textit{for all } c' \in C$$

$$\gamma = 1/f(c^t, d^t, c^{t+1}),$$

$$R_{ext}(c^{t+1}|c^t, d^t) \leftarrow R_{ext}(c^{t+1}|c^t, d^t) + \gamma \left(r_{ext} - R_{ext}(c^{t+1}|c^t, d^t)\right)$$

$$S_{ext}(c^{t+1}|c^t, d^t) \leftarrow S_{ext}(c^{t+1}|c^t, d^t) + \gamma \left(r^2_{ext} - S_{ext}(c^{t+1}|c^t, d^t)\right)$$

$$d_1 = \arg\max_{d \in D} \sum_{c'=1}^{|C|} R_{ext}(c'|c^t, d) \Pr(c'|c^t, d)$$

$$d_2 = \arg\max_{\substack{d \in D, \\ d \neq d_1}} \sum_{c'=1}^{|C|} R_{ext}(c'|c^t, d) \Pr(c'|c^t, d)$$

$$R_{mean}(c^t) = \frac{\sum_{d=d_1,d_2} \left[\left(\sum_{c'=1}^{|C|} f(c^t,d,c') \right) \left(\sum_{c'=1}^{|C|} R_{ext}(c'|c^t,d) \Pr(c'|c^t,d) \right) \right]}{\sum_{d=d_1,d_2} \left(\sum_{c'=1}^{|C|} f(c^t,d,c') \right)}$$

(STEP 4) Update the affect model:

$$\delta_E = R_{ext}(c^{t+1}|c^t, d^t) - R_{mean}(c^t)$$

$$\Pr(v'_0 = 1|c^{t+1}, c^t, a^t_0, d^t) \leftarrow \Pr(v'_0 = 1|c^{t+1}, c^t, a^t_0, d^t) - \epsilon\, \delta_E$$
$$\Pr(v'_0 = 2|c^{t+1}, c^t, a^t_0, d^t) \leftarrow \Pr(v'_0 = 2|c^{t+1}, c^t, a^t_0, d^t) + \epsilon\, \delta_E$$

(STEP 5) Update the uncertainty model:

$$\textit{For all } c \in C, \quad d \in D$$

$$\mu_0(c, d) = \sum_{c'=1}^{|C|} R_{ext}(c'|c, d) \Pr(c'|c, d)$$

$$\sigma_0(c, d) = \sqrt{\sum_{c'=1}^{|C|} S_{ext}(c'|c, d) \Pr(c'|c, d) - (\mu_0(c, d))^2}$$

(STEP 6) Update the extrinsic decision value:

$$
\begin{aligned}
&For\ all\ c \in C,\ \ d \in D \\
&Q_{ext}(c,d) := \sum_{c'=1}^{|C|} R_{ext}(c'|c,d)\Pr(c'|c,d) \\
&+ \alpha \sum_{c'=1}^{|C|} \max_{d'} Q_{ext}(c',d')\Pr(c'|c,d)
\end{aligned}
$$

(STEP 7) Get the new affective state a^{t+1}:

$$
\begin{aligned}
&a^{t+1} = (a_0^{t+1}, a_1^{t+1}, \cdots, a_{|A|}^{t+1}) \\
&For\ all\ \theta \in A = \{0,1,\cdots,|A|\} \\
&a_\theta^{t+1} = \{a_{\theta,1}^{t+1}, a_{\theta,2}^{t+1}\} : \\
&a_{\theta,1}^{t+1} = \Pr(v'_\theta = 1|c^{t+1}, c^t, a_\theta^t, d^t) \\
&a_{\theta,2}^{t+1} = \Pr(v'_\theta = 2|c^{t+1}, c^t, a_\theta^t, d^t)
\end{aligned}
$$

(STEP 8) For the next timestep,

$$
\begin{aligned}
&t \leftarrow t+1 \\
&\beta \leftarrow \beta + \beta_{inc}
\end{aligned}
$$

.....

This affective anticipatory reward model was partly inspired by the experiment by Bechara et al. [1] that made an experiment of the gambling task for two groups of people using the two types of decks of cards which have different reward distributions. Compared with patients with prefrontal damage, normals began to generate anticipatory skin conductance responses (SCRs) whenever they pondered a choice that turned out to be risky, before they knew explicitly that it was a risky choice, whereas patients never developed anticipatory SCRs although some eventually realized which choices were risky. In our model we assume that the normals' anticipatory responses are related to the uncertainty of choosing a decision and the responses can be interpreted as hope or uneasiness for a choice depending on the valence.

Now let's think of a special case in which only the seeking circuit is taken into account out of the affect systems and a new affective state for the seeking circuit a'_0 can be computed by the model $a'_0 = \Pr(v'_0|c', c, d)$ given a new cognitive state c', a current cognitive state c and a current decision d. In this case the following simpler version of the algorithm can be made: for (STEP 1),

$$
\Pr(d^t|c^t) = \frac{\exp(\beta Q_{DM}(c^t,d^t))}{\sum_{d=1}^{|D|}\exp(\beta Q_{DM}(c^t,d))}
$$

$$
\begin{aligned}
&For\ all\ d \in D \\
&Q_{DM}(c^t,d) := Q_{ext}(c^t,d) \\
&+ \eta_0\ \sigma_0(c^t,d) \sum_{c'=1}^{|C|}\sum_{v'_0=1}^{2} R_0(v'_0)\ \Pr(v'_0|c',c^t,d)\Pr(c'|c^t,d)
\end{aligned}
\tag{1}
$$

We assumed above that the affective anticipatory reward depended on the stochastic reward distribution of only the immediate extrinsic rewards without considering the delayed extrinsic rewards. However, in order to make a more efficient learning and decision making algorithm that utilizes the affective anticipatory reward to look for the optimal policy maximizing the long-run extrinsic rewards, we also suggest an algorithm in which we assume that the affective anticipatory reward for a choice depends on the distribution of the expected long-run extrinsic rewards for the choice.

For this new version, the following is applied for (STEP 4):

$$
\begin{aligned}
&V(c^{t+1}|c^t,d^t) = R_{ext}(c^{t+1}|c^t,d^t) + \alpha \max_{d'} Q_{ext}(c^{t+1},d') \\
&d_1 = \arg\max_{d\in D} \sum_{c'=1}^{|C|} V(c'|c^t,d)\Pr(c'|c^t,d) \\
&d_2 = \arg\max_{\substack{d\in D,\\ d\neq d_1}} \sum_{c'=1}^{|C|} V(c'|c^t,d)\Pr(c'|c^t,d) \\
&V_{mean}(c^t) \\
&= \frac{\sum_{d=d_1,d_2}\left[\left(\sum_{c'=1}^{|C|} f(c^t,d,c')\right)\left(\sum_{c'=1}^{|C|} V(c'|c^t,d)\Pr(c'|c^t,d)\right)\right]}{\sum_{d=d_1,d_2}\left(\sum_{c'=1}^{|C|} f(c^t,d,c')\right)} \\
&\delta_E = V(c^{t+1}|c^t,d^t) - V_{mean}(c^t)
\end{aligned}
\tag{2}
$$

where $V(c'|c,d)$ is the expected long-run rewards conditioned that the current cognitive state, the current decision and the next cognitive state are c, d and c', respectively.

In addition, whenever an episode ends, the following update rules are used for (STEP 5): the history of experience tuples for the episode is $\{< c^1, d^1, r_{ext}^1, c^2 > , < c^2, d^2, r_{ext}^2, c^3 >, \cdots, < c_n, d_n, r_{ext}^n, c^{n+1} >\}$,

$$
\begin{aligned}
&For\ k = 1,\cdots,n \\
&L(c^k,d^k) \leftarrow L(c^k,d^k) \\
&+(1/\sum_{c'=1}^{|C|} f(c^k,d^k,c'))\left((\sum_{j=k}^{n}\alpha^{j-k}r_{ext}^j) - L(c^k,d^k)\right) \\
&S(c^k,d^k) \leftarrow S(c^k,d^k) \\
&+(1/\sum_{c'=1}^{|C|} f(c^k,d^k,c'))\left((\sum_{j=k}^{n}\alpha^{j-k}r_{ext}^j)^2 - S(c^k,d^k)\right) \\
&For\ all\ \ c \in C,\ \ d \in D \\
&\mu_0(c,d) = L(c,d) \\
&\sigma_0(c,d) = \sqrt{S(c,d) - (\mu_0(c,d))^2}
\end{aligned}
\tag{3}
$$

where $L(c,d)$ and $S(c,d)$ give the expected values of long-run rewards and the square of long-run rewards, respectively, conditioned that a current cognitive state is c and a current decision is d, based on tuples from all the experienced episodes.

4 Experiments and Results

In order to evaluate the new model's ability to learn and make decisions, we performed several experiments. The ones reported here use the simpler version of the algorithm (Equation 1) for the convenience of computation.

We considered two kinds of stochastic reward distributions shown in Figure 1. Note that there are no

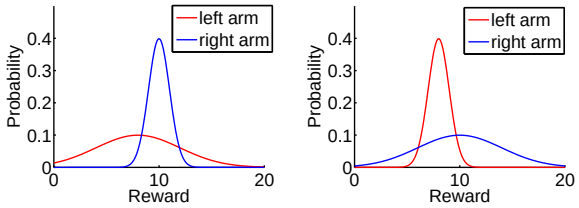

Figure 1: Two kinds of stochastic reward distributions of two arms: (a) right arm (mean = 10, std = 1), left arm (mean = 8, std = 4), (b) right arm (mean = 10, std = 4), left arm (mean = 8, std = 1).

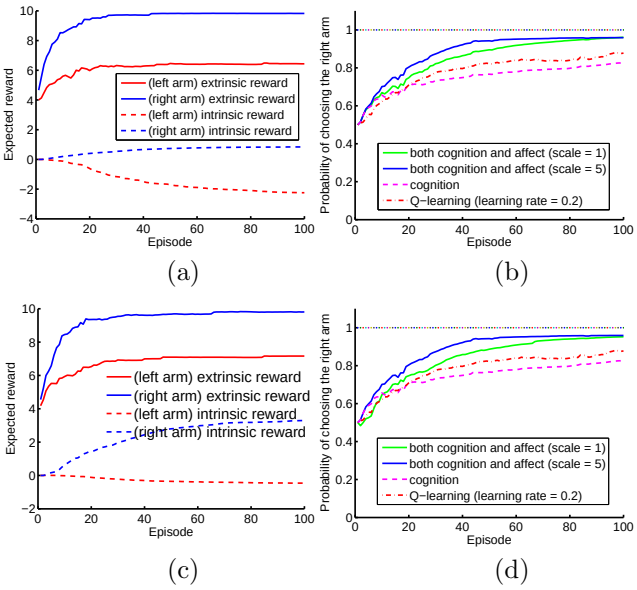

Figure 2: Single-step decision making (stochastic two-armed bandit type gambling tasks): the figures (a) and (b) for the reward distributions of Figure 1 (a), and the figures (c) and (d) for the reward distributions of Figure 1 (b). The lines show the averages over 100 experiments

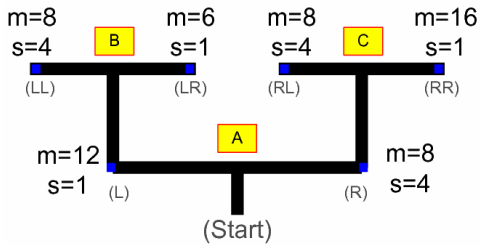

Figure 3: A maze for a sequential decision-making problem: A, B and C are the points for decision making. **m** and **s** denote the mean and the standard deviation of a reward distribution given at (L), (R), (LL), (LR), (RL) or (RR) points.

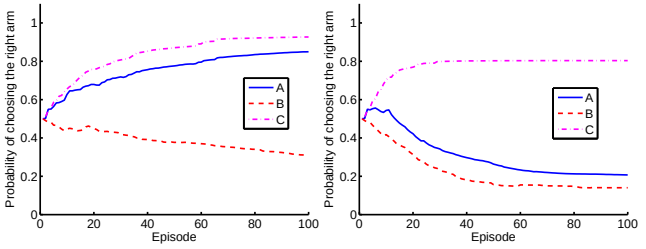

Figure 4: Sequential decision making (a maze task): the figures (a) and (b) show the probability of turning right for two different scale factors.

delayed future rewards for these kinds of single-step tasks.

In Figure 1 (a), the right arm has a reward distribution of the higher mean and the lower standard deviation, and the left arm has a reward distribution of the lower mean and the higher standard deviation. In this case, the optimal choice for a large number of times is the right arm because the right arm has the higher mean. Figure 2 (a) and (b) plots show the average results over 100 experiments for this case. Figure 2 (a) shows the expected extrinsic reward (Q_{ext}) and the expected intrinsic reward (affective anticipatory reward) for choosing each arm. Note that the expected extrinsic reward for choosing each arm converges to the mean of the corresponding reward distribution. Also, the magnitude of the affective anticipatory reward for choosing each arm is roughly proportional to the uncertainty or standard deviation of the corresponding reward distribution. Moreover, in this task the affective anticipatory reward for choosing the left arm is negative and greater in magnitude, but that for choosing the right arm is positive and smaller in magnitude. Since the mean of the reward distribution of the left arm is smaller than that of the right arm, in our affect model the valence for choosing the left arm is 'bad' (or negative) and the valence for choosing the right arm is 'good' (or positive). Thus, the uncertainty or arousal for choosing the left arm serves as a larger magnitude negative signal, which is analogous to uneasiness (or perhaps feeling risky), and that for choosing the right arm serves as a smaller magnitude but positive signal, analogous to a feeling of hope (or perhaps feeling lucky).

For the reward distributions of Figure 1 (b), Figure 2 (c) and (d) plots show the average results over 100 experiments. We can see from Figure 2 (c) that in this task the affective anticipatory reward for choosing the right arm is positive and greater in magnitude (as hope), but that for choosing the left arm is negative and smaller in magnitude (as uneasiness).

For both gambling tasks in Figure 1, we compared different algorithms: (i) both cognition and affect (using both extrinsic and intrinsic rewards) with $\eta_0 = 1$, (ii) both cognition and affect with $\eta_0 = 5$ (emphasizing intrinsic over extrinsic), (iii) cognition (using only

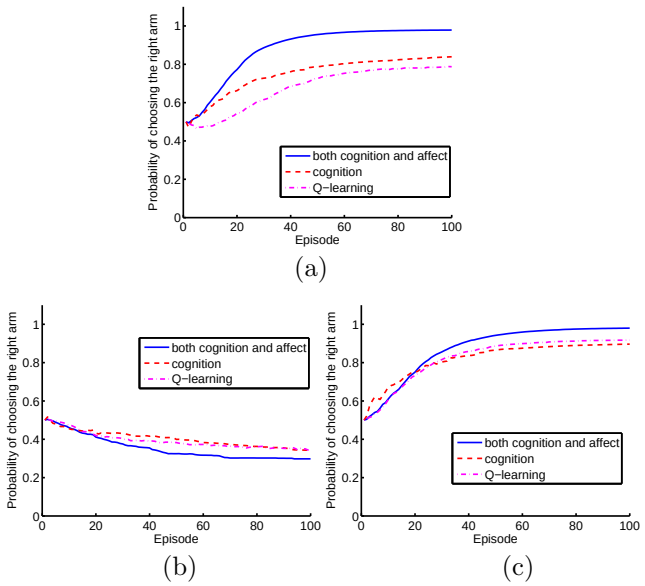

Figure 5: Comparison between three different algorithms: both cognition and affect (using Equations 2 and 3 with η_0=1), cognition (using only extrinsic rewards, η_0=0), and Q-learning (learning rate=0.2). The figures (a), (b) and (c), respectively, show the probability of turning right over episodes at the points A, B and C on the maze in Figure 3.

extrinsic rewards, $\eta_0 = 0$), and (iv) Q-learning with a constant learning rate of 0.2. All algorithms started with $\beta = 0.25$ and increased β by 0.005 each step. Also, $\epsilon = 0.05$ was used for updating the affect model. From Figure 2 (b) and (d), it is clear that the affective anticipatory rewards help faster learning of more accurate stochastic policies for these single-step decision making tasks. In other words, the anticipatory affective rewards help regulating the trade-off between exploration and exploitation. Although the algorithm starts with a smaller inverse temperature β facilitating exploration, affective biases result in facilitating exploitation as soon as a certain choice is found out to be a relatively higher mean.

In the case of sequential multi-step choices such as a maze task in Figure 3 where the delayed future rewards are involved in decision making, the immediate affective anticipatory reward can explain the well-studied phenomenon in economics of hyperbolic time discounting: people care more about the same time delay if it occurs earlier than if it occurs later [4]. That is, through a varying scale factor η_0 controlling the relative influence of the immediate affective anticipatory reward and the long-run rewards, a decision-maker (or an affective agent) can adjust the weight given to the immediate rewards over the future rewards in obtaining a decision making policy. As the scale factor η_0 is increased , the affective anticipatory reward makes the decision making give more weight on the short-term rewards. In Figure 4, we can see the difference between a small scale factor ($\eta_0 = 0.2$) and a large scale factor ($\eta_0 = 5$). In the case of a small scale factor (Figure 4 (a)), the affective agent learned a decision-making policy which weighted the future rewards as well as the immediate rewards, thus it turned right with a very high probability at the point A. However, in the case of a large scale factor (Figure 4 (b)), a decision-making policy weighted the immediate rewards more than the future rewards, thus the agent turned left with a very high probability at the point A.

Figure 5 (a), (b) and (c) plots, respectively, show the probability of turning right at the points A, B and C for different algorithms: (i) both cognition and affect (using Equations 2 and 3 with η_0=1), (ii) cognition (using only extrinsic rewards, η_0=0), and (iii) Q-learning with a constant learning rate of 0.2. Since we made use of Equations 2 and 3 in this case, the affective anticipatory reward depended on the distribution of the expected long-run rewards and it helped learning the optimal policy (that maximizes long-run rewards) faster than other algorithms.

5 Conclusions and Future work

We proposed a model of the affective anticipatory reward that is modeled by valence and arousal for a choice. Depending on whether valence is good or bad, the uncertainty can be interpreted as hope (feeling lucky) or uneasiness (feeling risky). Moreover, we showed that the integration of the affective anticipatory reward could be used for enhancing the efficacy of learning and decision making. Future work includes modeling of other affect systems in more detail and constructing a framework of enabling machines to make smarter and more human-like decisions for better human-machine interactions.

References

[1] Bechara, A., Damasio, H., Tranel, D., and Damasio, A. (1997). "Deciding Advantageously Before Knowing the Advantageous Strategy." *Science* 275: 1293-1295.

[2] Berridge, K. and Robinson, TE. (2003). "Parsing Reward." *Trends in Neurosciences* 26(9).

[3] Loewenstein, G. and Lerner J. S. (2003). "The Role of Affect in Decision Making." *Handbook of Affective Science*: 619-642.

[4] Loewenstein, G. and O'Donoghue, T. (2005). "Animal Spirits: Affective and Deliberative Processes in Economic Behavior."

[5] Moren, J. and Balkenius, C. (2000). "Reflections on Emotion." *Cybernetics and Systems* 2000.

[6] Oudeyer, P.-Y. and Kaplan, F. (2004). "Intelligent Adaptive Curiosity: A Source of Self-Development." *Proceedings of the 4th International Workshop on Epigenetic Robotics.*

[7] Panksepp, J. (1998). *Affective Neuroscience*, Oxford University Press.

[8] Singh, S., Barto, A. G., and Chentanez, N. (2004). "Intrinsically Motivated Reinforcement Learning." *Advances in Neural Information Processing* 18.