# Prediction in Human Decision Making: A Modeling Approach

**Rezvan Kianifar [1], Farzad Towhidkhah [1,*], Shahriar Gharibzadeh [1]**

1. Department of Biomedical Engineering, Amirkabir University of Technology, Iran.

Article info:
Received: June 20 2013
Accepted: March 04 2013

## ABSTRACT

Human beings can determine optimal behaviors, which depends on the ability to make planned and adaptive decisions. Decision making is defined as the ability to choose between different alternatives.

**Purpose:** this study, we have addressed the prediction aspect of human decision making from neurological, experimental and modeling points of view.

**Methods:** We used a predictive reinforcement learning framework to simulate the human decision making behavior, concentrating on the role of frontal brain regions which are responsible for predictive control of human behavior. The model was tested in a maze task and the human subjects were asked to do the same task. A group of six volunteers including three men and three women at the age of 23-26 participated in this experiment.

**Results:** The similarity between responses of the model and the human behavior was observed after varying the prediction horizons. We found that subjects with less risky choices usually decide based on considering long term advantages of their action selections, which is equal to the longer prediction horizon. However, they are more susceptible to reach suboptimal solutions if their predictions become wrong due to some reasons like changing environment or inaccurate models.

**Conclusion:** The concept of prediction result in faster learning and minimizing future losses in decision making problems. Since the problem solving in human beings is very faster than a trial and error system, considering this ability will help to describe the human behavior more desirably. This observation is compatible to the recent findings about the role of Dorsolateral Prefrontal Cortex in prediction and its relations to Anterior Cingulate Cortex with the ability of conflict monitoring and action selection.

**Keywords:**
Decision Making,
Prediction,
Maze Task,
Modeling.

## 1. Introduction

Decision making is defined as the ability to choose between different alternatives [1]. Regardless of how much the decision situation is complex, each decision making process includes three stages of forming preferences, selecting and executing actions, and evaluating the outcomes [2].

Today, functional magnetic resonance imaging (fMRI) technique is the most popular method used for the investigation of cognitive brain functions [3, 4, 5]. For this purpose, some cognitive tasks should be designed, under which the behavior of the subjects and their brain activity is being observed.

There are a huge number of studies which have used these tasks together with statistical methods to define the involvement of the specific brain areas in various cognitive tasks [3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. These studies have reported some patterns of brain activity, based on which there are many hypotheses about the function of various brain regions in higher order tasks like decision making.

*** Corresponding Author:***
Farzad Towhidkhah, PhD
Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran..
Tel: +98 21 6454 2363
E-mail: Towhidkhah@aut.ac.ir / Towhidkhah@gmail.com

Moreover, some attempts of the recent years are dedicated to the application of theoretical methods for modeling, explanation and prediction of biological mechanisms. In this regard, it has become very useful to explore the ideas from reinforcement learning theory to the psychology of reward-based cognitive functions because of the established evidences on existing many common aspects between them [15-26]. The development of biologically inspired models could have many advantages in studying cognitive functions of the brain as they do not have the same limitations as the study of real subjects do.

One of the important aspects of human decision making refers to the prediction and anticipatory capabilities [27]. Although there have been many studies focusing on the role of prediction in decision making, most of which are limited to the neural level of only predicting the reward signal, for example a few seconds before its delivery [11, 13, 14, 28, 29].

In this study, we have addressed the prediction aspect of human decision making from neurological, experimental and modeling points of view. Although our study is not limited to the neural level, we have considered the brain regions involved in prediction as modules with a defined function and investigated the overall effect of interaction between these regions on human decision making behavior.

For this purpose, we first introduced the most important brain regions involved in prediction and, then, proposed the appropriate reinforcement learning (RL) architecture to model this process. Finally, we applied a maze task to simulate the introduced model, and designed the same environment for human participants to have an index to evaluate the model's function.

## 2. Methods

### 2.1. Neurological Aspect

Today, it is revealed that human beings employ a reinforcement learning process to decide between alternative options [14, 28]. Several cortical and sub-cortical regions are involved in a decision making process, most of which have a reward related activity [13].

Although a great number of brain regions may be involved in reward processing, the activity of Amygdala, Striatum and the Prefrontal, Orbitofrontal and Anterior Cingulate Cortexes are reported in most of the recent studies [2, 3, 5-7, 10, 12, 30-38].

Coricelli and his colleagues have distinguished two levels of reward processing in the brain [31]. First-level processing is based on signaling within dopaminergic neurons in which the brain is not able to discriminate between different rewards (alternatives). Reward processing in the second level is related to the neuronal activity in regions such as the Orbitofrontal Cortex (OFC), Anterior Cingulate Cortex (ACC), and, perhaps, the Amygdale [31].

In the present study, we will focus on the second level and the way it could provide us with proper predictions of future events.

### 2.1.1 Amygdale and Reward Production

Amygdale is a neural substrate involved in the motivational aspect of decisions [3, 39]. This region encodes external or internal desirability of actions in terms of rewards. External reward is given by the environment as a result of doing actions while internal reward refers to the self-satisfaction of doing a special action which is related to the emotional content of that action. Amygdale also involves in emotional processing of events [2, 31].

### 2.1.2. DorsoLateral PreFrontal Cortex and future Prediction

DorsoLateral PreFrontal Cortex (DLPFC) is known as a neural substrate for working memory, activation of which have been reported in most of decision making tasks [1, 2, 13, 29, 35]. Working memory refers to the ability of DLPFC for transient representing, maintaining and manipulating of task-relevant information which is not immediately present in the environment [40, 41].

Based on this ability, DLPFC is a substrate for model-based processes and is able to predict future state and reward expectancy in a predictable environment [4, 29]. In other words, a temporal model of the current decision state will be formed in the working memory, using which the DLPFC is able to predict future states and plan a sequence of future actions.

### 2.1.3. Anterior Cingulate Cortex and Action Selection

The Anterior Cingulate Cortex has a central role in the action selection and making voluntary choices [6, 10, 33, 35, 36]. ACC is closely connected to the motor system and its lesions impair reinforcement-guided action selection [24, 29].

The randomness in action selection, which should be controlled during the learning process, is called exploration-exploitation problem. It is suggested that ACC can generate these exploratory actions in order to control the exploration in a novel or changed environment [33].

ACC has also been suggested as a monitoring center for detecting and managing response conflicts which could arise between mappings of stimuli-response in the working memory [35].

Cohen and his colleagues have confirmed that ACC engages in behavioral control, error monitoring, and reward calculations [6].

A recent review has suggested that dorsal anterior cingulate cortex (dACC) specifies control signals that maximizes estimated expected value of control which, then, will be implemented to make changes in information processing in other regions of the brain to perform a specified task [42].

### 2.1.4. OrbitoFrontal Cortex and Action Evaluation

OrbitoFrontal Cortex (OFC) plays a critical role in decision making process [6, 11, 31, 33]. Correlations have been found between the magnitude of outcomes and the OFC's activity [11].

Patients with lesions to OFC incline to prefer risky decision strategies regardless of the long-term outcomes of those strategies. OFC removal in rats disables them to guide their behavior based on the reward evaluation [33]. OFC can also integrate cognitive and emotional information due to its interactions with Amygdale [31].

ACC and OFC receive similar reward information through connections with the Ventral Striatum and Amygdale. However, OFC has relatively greater links to the stimuli information while the ACC is more bounded to the spatial and motor systems [33]. The probable role of the OFC is, then, to update representation of values and bias responses according to their relative reward value [6].

### 2.1.5. Striatum and Error Prediction

It is well demonstrated that Dopaminergic (DA) neurons (especially in Ventral Striatum) engage in processing reward stimuli [3, 31, 33, 34, 36]. DA neurons are activated by the stimuli associated with reward prediction, and it is suggested that these neurons represent error in the prediction of future reward [1, 2, 13, 38].

Tanaka and his colleagues suggest that the OFC-ventral striatum loop is involved in action learning based on the present state, while the DLPFC-dorsal striatum loop acts based on the predictable future states [13].

### 2.1.6. Functional Relationships

According to the mentioned findings, we can extract the following functional relationships between various brain regions during a predictable decision making task:

- Sensory information together with other information about the current decision state are recalled from related brain areas into the DLPFC where a temporal model of environment will be formed.
- DLPFC provides proper predictions about the future states and rewards based on this model. Because of the limited space of working memory, a limited number of states could be processed simultaneously at each decision step. We have interpreted these limited states as prediction horizon which will be updated after transition to the next step.
- The desirability of each action selection is defined by means of rewards. Amygdale is a substrate which provides us with reward signals in accordance with the external or internal emotional content of actions.
- In the Ventral Striatum, Dopaminergic (DA) neurons provide prediction error which occurs during the processing of reward signal. OFC has access to the information about reward, prediction error and the predicted future states, enabling it to evaluate the predicted future states [33].
- Finally, ACC will select the appropriate actions based on OFC evaluations to be sent to the motor units of the brain. At the same time, adjustments to the model will be done in DLPFC if conflict information is reported from ACC. This conflict can be detected by comparing the acquired accumulated rewards with those expected according to the model.

After several times of successful selection of a sequence of actions which have always guided us to the maximum reward, those actions will be considered as a unique option. It means that when exposed to the first state of a sequence, all of the actions of that sequence will be selected automatically one after another without any need to search for other optimal actions. This is the concept of hierarchy in our decisions which is fully described by Botvinick et al. [15, 26]. Options are valid as long as they provide us with the maximum expected reward. The benefit of forming options is to speed up the decision making process through avoiding redundant repeats. Performing of options will be done by a lower cognitive level of the brain hierarchy. Basal ganglia is suggested as a neural substrate where options will be formed and controlled [22].

Although this hierarchy has a critical role in our responses , we have only concentrated on the first (conscious) level of decision making in the present study, emphasizing the role of predictions in a decision making process. Therefore, we will not insert any option learning in the model. Figure 1 is a schematic diagram of the above described interactions.

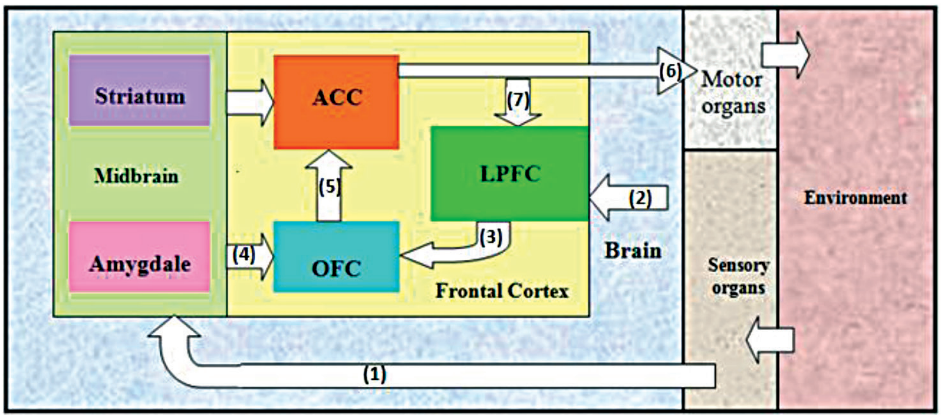


**Figure 1.** The relationship between different brain areas in a predictable decision making task. Arrows represent the information flow. Sensory information is received in the midbrain which results in producing reward and reward prediction error signals in Amygdale and Striatum, respectively (1). Based on the information recalled from related brain regions (2), a model of the current decision step will be formed in DLPFC using which DLPFC provides prediction about future states (3). The OFC has access to the information about the reward, prediction error (4) and the predicted future states (3). It can evaluate predictions by means of error information. Finally, ACC will select appropriate actions based on OFC evaluations (5) to be sent to the motor units of the brain (6). At the same time, adjustments to the model will be done in the DLPFC if conflict information is reported through ACC (7).

## 2.2. Modeling of the Relationships

Reinforcement learning theory is known as a powerful method to explain reward-based activity of the brain. Many similarities have been suggested between the elements of this method and the related neural mechanisms [15, 16, 21, 25].

RL addresses the interaction between an agent and the environment. RL schemes are formulated in terms of Markov decision process (MDP) as an optimal control problem [43].

The main elements of MDP are state (s), action (a), reward (r), transition probability P(s'|s,a), and policy (π) [43]. Reward is a scalar value which indicates instantaneous goodness of an action. The objective of the agent is to maximize accumulated rewards toward the future, which is done by improving its strategy. Such strategy is called a policy which is a mapping from each state and action to the probability of choosing that action. Estimation of the reward accumulation is called the value function V(s).

P(s'|s,a) is the probability of reaching state (s') by selecting an action (a) at state (s), which is usually unknown. Temporal-deference (TD) learning methods are the reinforcement learning solution algorithms which try to approximate the value function based on the agent's experience without any requirement to know the transition probabilities. We have used an Actor-Critic architecture for solving the RL problem.

Actor-Critic is a TD method with a separate memory structure to represent the policy independent of the value function [43]. The Actor selects actions according to the policy based on a set of weighted associations from states to actions called action strengths. The Critic maintains a value function associating each state with an estimate of an accumulated reward expected from that state. Both the action strength and the value function must be learned empirically. At the beginning of training, action strengths and state values are initialized to zero. In our task, actions are deterministic which means the transition probability for each state and chosen action is one, but it does not impose any limitation to the model.

Actor-Critic implementation is done according to Botvinick et al. formulation [15]. The Actor includes a matrix of real-valued strengths (W) for each action in each state. The Critic maintained a vector V of values, attaching a real number to each state. Action strengths (W) and value function (V) will be updated using temporal-difference (TD) prediction error as shown in Equation (1):

$$\begin{aligned} &V(s_t) \leftarrow V(s_t) + \alpha_C \delta \\ &W(s_t, a) \leftarrow W(s_t, a) + \alpha_A \delta \end{aligned} \tag{1}$$

TD- prediction error (δ) is computed according to Equation (2):

$$\delta = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \tag{2}$$

Where γ is a discount factor and $\alpha_C$, $\alpha_A$ are the learning rate parameters. $S_{t+1}$ is the next state and $r_{t+1}$ is the acquired reward related to the next state.

A positive prediction error will increase the value of the previous state and the tendency of reselecting the chosen action at that state.

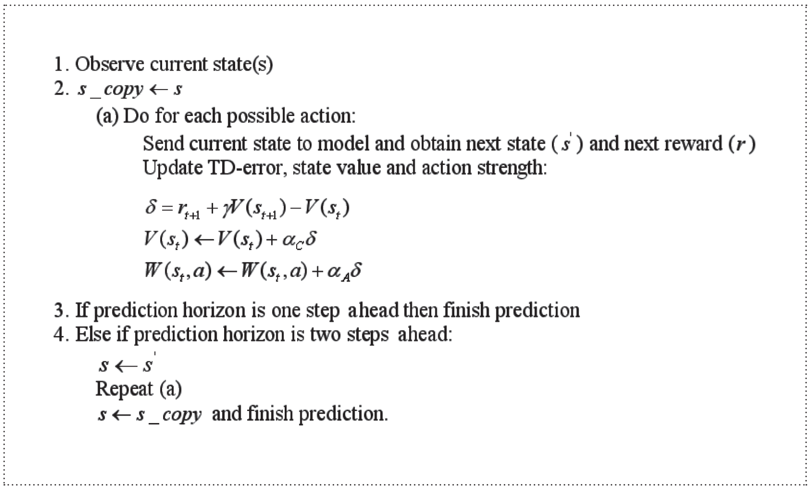
Action selection is according to the Softmax equation (Equation (3)):

$$P(a) = \frac{e^{W(s_t,a)/\tau}}{\sum_{a' \in A} e^{W(s_t,a')/\tau}} \tag{3}$$

Where P(a) is the probability of selecting action (a) at step (t); $W(s_t,a)$ is the weight for action (a) in the current state; τ is a temperature parameter which controls the exploration-exploitation tendency; and A is a set of all actions [15].

We have added a prediction part to the Actor-Critic structure to address a model of human decision making process which includes working memory. For this purpose, it is necessary to have an internal model of the external environment like all other predictive structures. Given the current state and action, this model will return next state and corresponding reward [43].

Agent begins from a start state, and recalls a part of the model which is related to the current state before choosing any action. Then, agent estimates the value of going to each of the neighboring states using an available part of the model and updates the value function of the related neighbors. How far from the current state one can evaluate, depends on the prediction horizon. We have considered one and two steps ahead prediction horizons. Figure 2 shows the algorithm used to predict future states.



```
1. Observe current state(s)
2. $s\_copy \leftarrow s$
    (a) Do for each possible action:
        Send current state to model and obtain next state ($s'$) and next reward ($r$)
        Update TD-error, state value and action strength:
        $\delta = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$
        $V(s_t) \leftarrow V(s_t) + \alpha_C \delta$
        $W(s_t, a) \leftarrow W(s_t, a) + \alpha_A \delta$
3. If prediction horizon is one step ahead then finish prediction
4. Else if prediction horizon is two steps ahead:
    $s \leftarrow s'$
    Repeat (a)
    $s \leftarrow s\_copy$ and finish prediction.
```

**Figure 2.** The prediction algorithm.

Considering the neural mechanisms introduced in section 2.1 and the introduced RL structure, we can extract the following similarities between them:

ACC involves in the action selection so it could be modeled as Actor which includes a matrix of strength weights. The goodness of the transition to each state is evaluated by the Critic which has the same role as OFC does in neural system.

Both state value and action strength are updated by means of prediction (TD) error which is produced by

Dopamine neurons. Moreover, serotonergic system controls the time scale of evaluation (γ) and Acetyl cholinergic system controls the learning rate (α) [16].

Working memory (DLPFC) has a limited capacity in number of states recalling simultaneously which is equal to the prediction horizon in the RL structure. Therefore, the proposed algorithm for prediction is similar to that of working memory. In simulations, we have assumed that a model of environment is available, which corresponds to the internal model in working memory, but agent could have access to only a part of it at each decision step.

Amygdale involves in representation of rewards. Here, we have considered only the external rewards. This region is modeled using a matrix of scalar values which represents the reward amount for each state of the environment.

The proposed model for decision making process together with its RL realization is shown in Fig. 3.

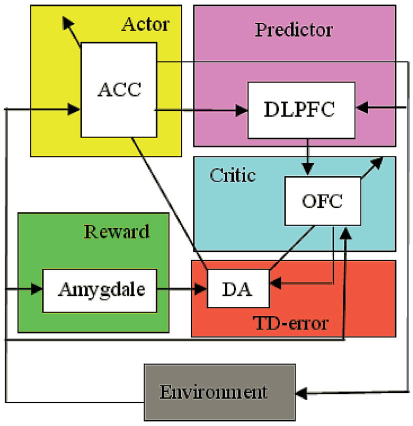


**Figure 3.** The RL realization of the functional model of Figure 1.

## 2.3. Simulation Task

To examine the model, we have chosen two maze tasks shown in Fig. 4 (a and b), in which RL agent has to go from the start state to the goal state taking a minimum number of steps. There are four actions, up, down, right and left in each state which take the agent deterministically to the corresponding neighboring states, except when movement is blocked by an obstacle or the edge of the maze in which case the agent remains where it is and receives the reward of -1. Reward is zero on other transitions except into the goal, on which it is +1. We have run the simulations under three conditions: when there is no prediction horizon, one step ahead prediction horizon and two steps ahead prediction horizon (Fig. 4 (c and d)). The two steps ahead prediction horizon includes those states which could be achieved by performing two consequent allowed actions. Therefore, this horizon does not have the shape of a complete square. The location of start and goal states and the pattern of the mazes are similar for both horizons.

The results of the simulations are shown in Fig. 5. It shows the learning curves of the three mentioned conditions. These curves are the number of steps to goal per episode which are averaged over 10 runs for each condition.

## 2.4. Experimental Task

We put the same mazes of the simulation task into an experiment to evaluate the behavior of human subjects. The environment is designed utilizing Matlab (GUI) software as shown in Fig. 6.

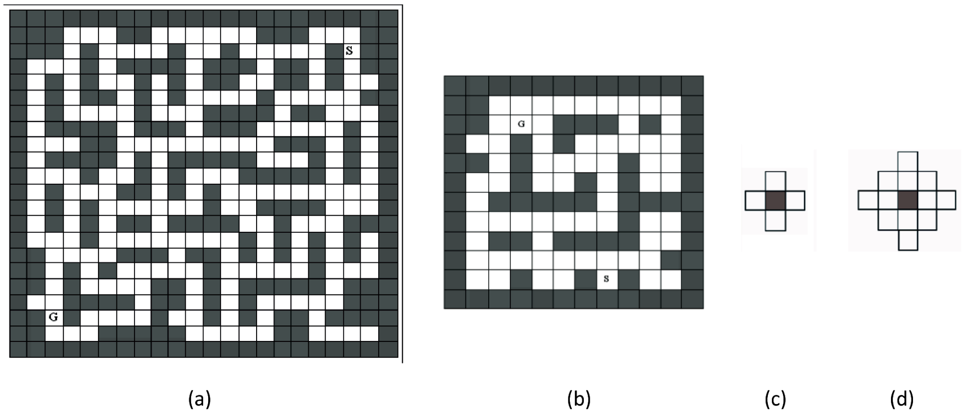


**Figure 4. a, b)** the environments of the simulation and the experimental tasks. S and G refer to the start and goal states, respectively. **c)** One step ahead prediction horizon considering that the dark state is the current state **d)** two steps ahead prediction horizon.

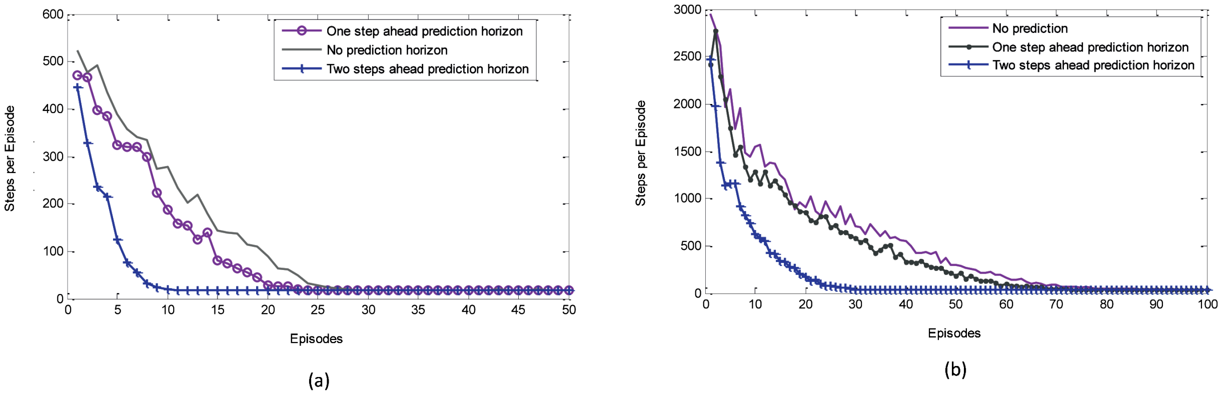
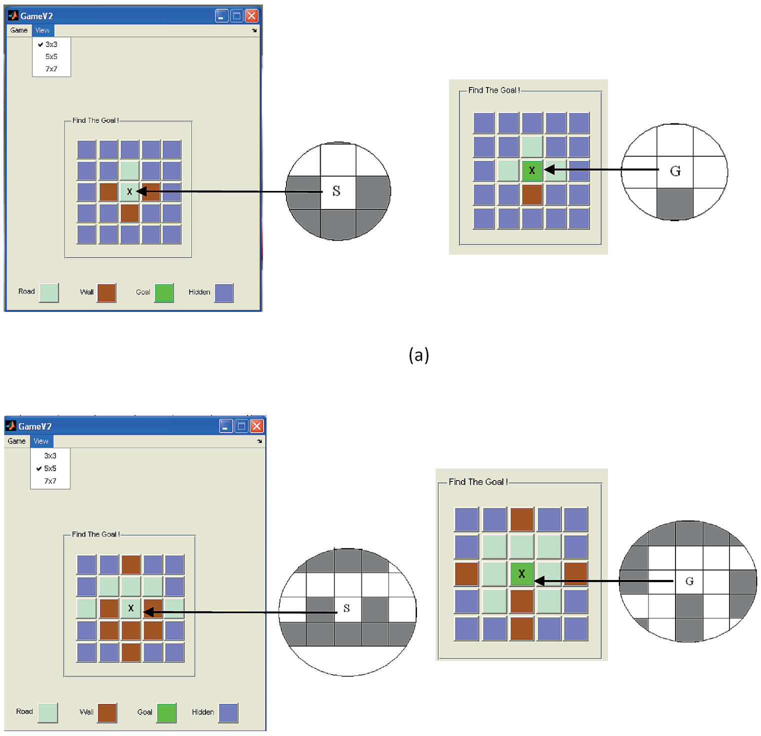


**Figure 5.** Number of steps per episode for three different conditions of a task in the small (a) and big (b) mazes: when there is no prediction horizon, one step ahead prediction horizon and two steps ahead prediction horizon.



(a)

(b)

**Figure 6.** The designed environment for experimental task. In this environment, multiplication sign refers to current state. The goal state is shown in green, the walls are brown and the paths are light blue. The dark blue states are hidden states which cannot be seen. In each decision state, a limited space of the maze is visible which represents the prediction horizon. This view gets updated after going to the new state by clicking on it. Newly selected state could be one of the four adjacent states in north, east, south and west of the current state, provided that these states are not walls. No transition will take place by clicking on the walls, hidden states or states which are not adjacent to the current state. For the purpose of this experiment two horizons are available: one step ahead (a) and two steps ahead (b), which can be chosen from the view menu.

In this environment, the current state is defined by multiplication sign. The goal state, walls and paths are shown in green, brown and light blue, respectively. The states shown in dark blue are hidden states. Transition to the next chosen state is done by clicking on it. Player is allowed to go to the four adjacent states in north, south, east and west of the current state. Other states are inactive so that no transition will take place by clicking on them. A limited view around the current location is visible which is called prediction horizon. This horizon gets updated after going to the new state. For the purpose of this experiment, two horizons are available: one step ahead horizon and two steps ahead horizon which can be selected before the beginning of the task from the view menu.

Before beginning the experiment, subjects are asked to move through the environment to get familiar with the moving procedures. When it is confirmed that they have learned how to move, the task starts.

At the beginning, subjects are placed in the start state and try to reach the goal in a trial and error manner. They are not aware of the location of the start and goal states and the shape of the maze. They are only informed about the number of minimum steps from start to the goal. After reaching the goal, all of their visited states will be saved in an excel file, and they begin the next episode with the same start and goal positions. After a few repeats, they can estimate the goal location in relatation to themselves and, then, with respect to the visible horizon, they can predict how close they are to the goal which helps them to choose the correct direction.

Subjects will repeat the same episode until they reach the goal in minimum steps. The number of repeats depends on the learning and prediction ability of the subjects. After finding the optimum path, they are asked to repeat it for two more times, which ensure us that the optimum solution is not achieved by accident.

## 3. Results

A group of six volunteers including three men and three women at the age of 23-26 participated in this experiment. Participants were recruited via announcements on the Amirkabir University of Technology campus. Each of the subjects did four tasks including two mazes with two different horizons. Since the positions of the start and goal states remain unchanged as the horizon increases, the tests were taken in two days. In the first day, they found the goal in two mazes with one step ahead horizon and in the second day with two steps ahead horizon. Meanwhile, they are not told that they are experimenting the same mazes as the previous tasks. This will ensure us that the only parameter which affects the behavior of the subjects is the changing of the prediction horizon.

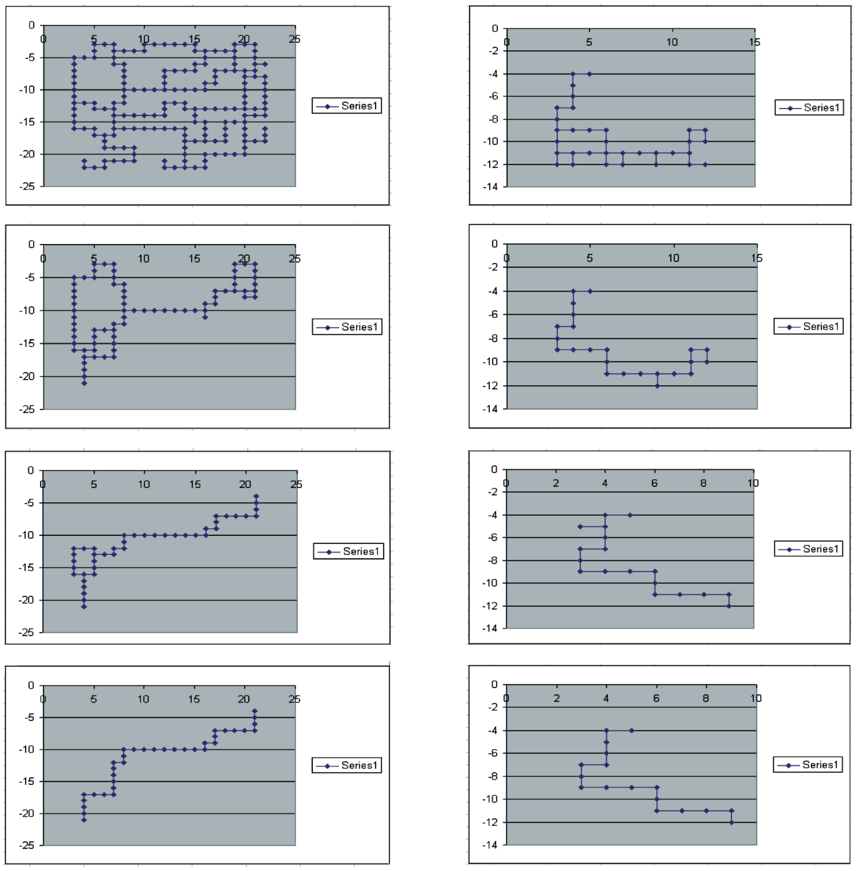


**Figure 7.** The visited states and acquired paths in four consequent episodes for one of the participants. Gray background represents maze environment and blue dots and lines refer to the visited states and paths, respectively. Figures in the left column relate to the big maze while figures in the right column are those for the small maze. Panels from top to bottom of each column represent the visited states and acquired paths in the first, second, third and last episodes, respectively. The paths of the bottom panels are optimal paths with minimum number of visited steps.

Fig. 7 shows visited states and acquired paths in four consequent episodes for one of the participants. Figures in left column relate to the big maze and figures in right column are those for the small maze. Panels from top to bottom represent the acquired paths in the first, second, third and last episodes, respectively. It can be seen that the number of visited states are reduced from top to bottom panel, and the shape of paths get closed to the optimal path which shows the learning process. The paths of the bottom panels are optimal paths with a minimum number of visited steps.

Learning curves (number of steps per episode) of the sixth participants are shown in Fig. 8. Left and right columns relate to the small and big mazes, respectively. Panels depicted in each row relate to the same subject. Each panel includes two graphs which are the learning curves of the same subject with one step ahead (distinguished with circles) and two steps ahead (simple line) prediction horizons.

## 4. Discussion

With respect to the results shown in Fig. 8, it is evident that the more the prediction horizon increases, the faster the convergence rate of the responses becomes. It means that with a bigger horizon, the optimum path is found more quickly. That is because of the fact that with a bigger horizon, we can process more information in our working memory so that we will have a better estimation of where we are and what the possible position of the goal in relation to our current place would be. This result was also confirmed in our simulations (Fig. 5).

Moreover, by comparing three top graphs of Figure 8 with three bottom ones for each maze, two decision making strategies are detectable. For the sake of a better comparison, we have depicted the learning curves of all of the subjects for the small maze in Figure 9.

According to the experimental results, some of the subjects have a slope response while others have a flat one. Subjects with the slope response are those who persist on exploiting their previous experiences. Therefore, their decisions are less risky. They could better remember their previously passed states so that they can construct a better internal model of the environment in their working memory, and have a longer prediction horizon. Due to this bigger prediction horizon, they can better estimate the approximate location of the goal in relation to themselves after one or two episodes. In the next re-

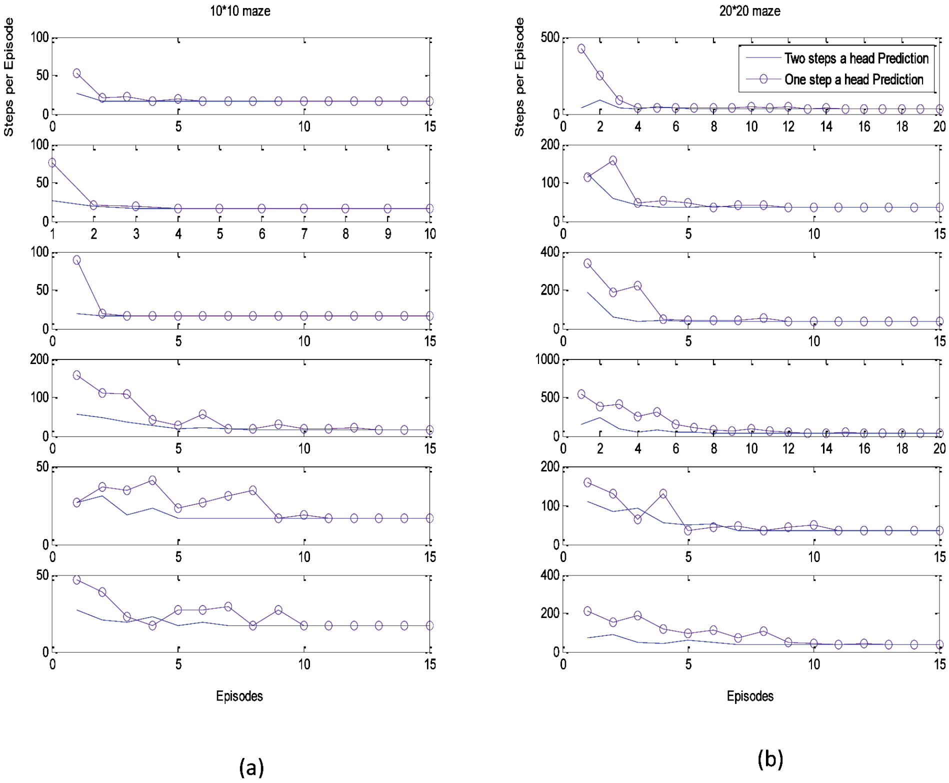


**Figure 8.** The number of steps per episode for experimental tasks. Column (a) is for the small maze and column (b) is for the big maze. The panels in the same row are related to the same subject. Each panel includes two graphs which are the learning curves of the same subject with one and two steps ahead prediction horizons.

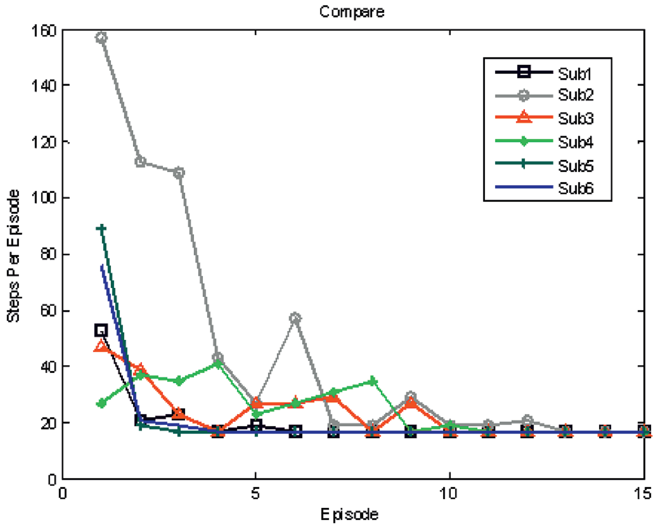


**Figure 9.** The learning curves of all of the sixth subjects for the small maze and one step ahead prediction horizon.

peats, they only try to shorten their previous path using their predictions.

Although the mentioned group can act faster than the group with a flat response, they are more susceptible to reach the suboptimal solutions. In this task, if their selected states in the first episodes do not include the shortest path, their internal model will be incomplete. They will never find the shortest path since they choose actions more based on this model and less randomly. An example of this sub-optimality is seen in Fig. 10 which shows the chosen paths of one of the mentioned subjects in the small maze.

The shortest path of this maze includes 17 states. This subject has reached the goal from the right side of the maze for the first time (as shown in panel (a) of Fig. 10) while the shortest path is through the left side. He has tried to optimize the same path in his next attempts (panel (b, c) of Fig. 10) so he could not find the shortest path at the end of his trials (panel d of Fig. 10).

On the other hand, the second group of participants are those who like to take riskier actions. They have a smaller prediction horizon but they are still less prone to reach sub-optimal solutions.

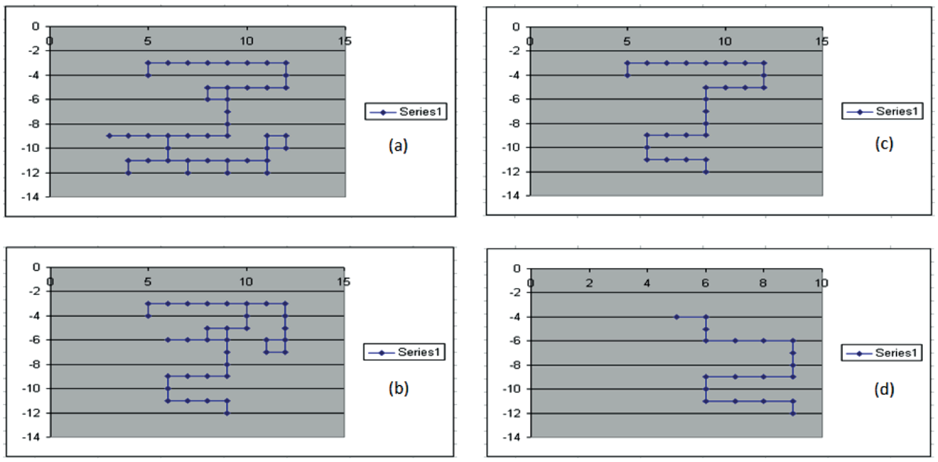


**Figure 10.** The chosen paths by one of the subjects in the small maze. Gray background represents maze environment. In each panel, blue dots are visited states while blue line shows acquired path. Panels (a, b, c, d) represent passed states in the first, second, and third episodes, respectively. Panel (d) relates to the last episode (fourth) which is a sub optimal solution.

Some of the previous studies have modeled these two kinds of decision strategies using different probabilities in action selection [17, 18, 19]. However, in a predictable environment (which means that people have learnt it before), human subjects always rely on their previous knowledge of that environment (internal model) unless they understand that their knowledge is no more valid. Therefore, we believe that this observation could be modeled more efficiently using various prediction horizon concepts because it is more compatible to the recent findings about the role of DLPFC in prediction and its relations to ACC with the ability of conflict monitoring.

In conclusion, the concept of prediction will result in faster learning and minimizing future losses in decision making problems. Since the problem solving in human beings is very faster than a trial and error system, considering this ability will help to describe the human behavior more desirably.

Although our model acted like the human subjects qualitatively, it is not still a complete one because of the quantitative differences between the results. We believe that another important ability which has to be considered for a more complete modeling is the concept of hierarchy in decision making process because it will help us to do well learned tasks automatically and concentrate on solving the new problems which results in very fast responses.

In the future works, we can design some experiments to show the importance of hierarchy and ,then, add it to the model. Also, we can investigate the adaptation ability in decision makings which is more related to the function of ACC in conflict detection and updating of the internal model. This is especially important when we have a dynamic environment.

Moreover, the tasks can be done on patients who suffer from decision making deficits. Then, we can try to simulate their results by varying the parameters of the proposed model. This will help us to follow which region is more susceptible to damage in each case. A good candidate of patients for this could be those with ventromedial prefrontal cortex lesions who suffer from myopia in decision making [44].

Finally, it is remarked that the proposed model and experiments design in this study are new and due to limited number of the participants (similar to [7, 8, 17, 23]), the results show primarily founding; further experiments with larger number of subjects are required for better results validation.

## References

[1] Braclough, D. J., Conroy, M. L., & Lee, D. (2004). Prefrontal cortex and decision making in mixed-strategy game. Nature Neuroscience, 7(4), 404-410.

[2] Ernst, M., & Paulus, M. P. (2005). Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. Biological Psychiatry, 58, 597-604.

[3] Yarkoni, T., Gray, J. R., Chrastil, E. R., Barch, D. M., Green, L., & Braver, T. S. (2005). Sustained neural activity associated with cognitive control during temporally extended decision making. Cognitive Brain Research, 23, 71-84.

[4] Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of Dorsolateral Prefrontal Cortex decreases model-based in favor of model-free control in humans. Neuron, 80, 914–919.

[5] Whitman, J. C., Metzak, P. D., Lavigne, K. M., & Woodward, T. S. (2013). Functional connectivity in a frontoparietal network involving the dorsal anterior cingulate cortex underlies decisions to accept a hypothesis. Neuropsychologia, 51, 1132-1141.

[6] Cohen, M. X., Heller, A. S., & Ranganath, C. (2005). Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. Cognitive Brain Research, 23(1), 61-70.

[7] Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. Journal of Neuroscience, 20 (16), 6159-6165.

[8 ] Ernst, M., Kimes, A. S., London, E. D., Matochik, J. A., Eldreth, D., Tata, S., et al. (2003). Neural substrates of decision making in adults with attention deficit hyperactivity disorder. American Journal of Psychiatry, 160, 1061–1070.

[9] Hampton, A. N., Bossaerts, P., & O' Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. Journal of Neuroscience, 26(32), 8360-8367.

[10] Paulus, M. P., & Frank, L. R. (2006). Anterior cingulate activity modulates nonlinear decision weight function of uncertain prospects. NeuroImage, 30(2), 668-677.

[11] Polezzi, D., Lotto, L., Daum, I., Sartori, G., & Rumiati, R. (2008). Predicting outcomes of decisions in the brain. Behavioral Brain Research, 187(1), 116-122.

[12] Smith, B. W., Mitchell, D. G. V., Hardin, M. G., Jazbec, S., Fridberg, D., Blair, R. J. R., et al. (2009). Neural substrates of reward magnitude, probability,and risk during a wheel of fortune decision-making task. NeuroImage, 44(2), 600-609.

[13] Tanaka, S. C., Samejima, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., et al. (2006). Brain mechanism of reward prediction under predictable and unpredictable environment dynamics. Neural Networks, 19(8), 1233-1241.

[14] Silvetti, M., Castellar, E. N., Roger, C., & Verguts, T. (2014). Reward expectation and prediction error in human medial frontal cortex: An EEG study. Neuroimage, 84, 376-382.

[15] Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations:A reinforcement learning perspective. Cognition, 113, 262-280.

[16] Doya, k. (2002). Metalearning and neuromodulation. Neural Networks, 15(4-6), 495-506.

[17] Ishida, F., Sasaki, T., Sakaguchi, Y., & Shimai, H. (2009). Reinforcement-learning agents with different temperature parameters explain the variety of human action-selection behavior in a markov decision process task . Neurocomputing, 72, 1979-1984.

[18] Ishii, S., Yoshida, W., & Yoshimoto, J. (2002). Control of exploitation- exploration meta-parameter in reinforcement learning. Neural Networks, 15, 665-687.

[19] Kalidindi, K., & Bowman, H. (2007). Using ε-greedy reinforcement learning methods to further understand ventromedial prefrontal patients' deficits on the Iowa Gambling Task. Neural Networks, 20, 676-689.

[20] Kawato, M., & Samejima, K. (2007). Efficient reinforcement learning: Computational theories, neuroscience and robotics. Current Opinion in Neurobiology, 17(2), 205-212.

[21] Montague, R. P., Eagleman, D. A., McClure, S. M., & Berns, G. S. (2006). Reinforcement learning: A biological perspective. Encyclopedia of Cognitive Science. John Wiley & sons, Ltd.

[22] Pisapia, N. D. (2004). A framework for implicit planning, towards a cognitive/computational neuroscience theory of prefrontal cortex function. PhD. Thesis, Edinburge University.

[23] Stankiewicz, B. J., Legge, G. E., Mansfield, J. S., & Schlicht, E. J. (2006). Lost in virtual space: Studies in human and ideal spatial navigation. Journal of Experimental Psychology:Human Perception and Performance, 32(3), 688-704.

[24] Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2013). From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

[25] Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. Psychological Review, 119(1), 120-154.

[26] Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. Current Opinion in Neurobiology, 22, 956–962.

[27] Butz, M. (2004). Anticipation for learning, cognition, and education. On the Horizon, 12, 111-116.

[28] Cohen, X. M., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. Journal of Neuroscience, 27(2), 371-378.

[29] Tsujimoto, S., & Sawaguchi, T. (2005). Neural activity representing temporal prediction of reward in the primate prefrontal cortex. Journal of Neurophysiology, 93(6) 3687-3692.

[30] Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behaviour. Trends in Cognitive Sciences, 11(3),118–125.

[31] Coricelli, G., Dolan, R. J., & Sirigu, A. (2007). Brain, emotion and decision making: The paradigmatic example of regret. Trends in Cognitive Sciences, 11(6), 258-265.

[32] Pais-Viera, M., Lima, D., & Galhardo, V. (2007). Orbitofrontal cortex lesions disrupt risk assessment in a novel serial decision making task for rats. Neuroscience, 145(1), 225-231.

[33] Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends in Cognitive Sciences, 11(4), 168-176.

[34] Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. Science, 318(5850), 598-602.

[35] Sohn, M. H., Albert, M. V., Jung, K., Carter, C. S., & Anderson, J. R. (2007). Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. Proc Natl Acad Sci USA., 104(25), 10330-10334.

[36] Walton, M. E., Croxson, P. L., Behrens, T. E. J., Kennerley, S. W., & Rushworth, M. F. S. (2007). Adaptive decision making and value in the anterior cingulate cortex. NeuroImage, 36, T142-T154.

[37] Coutlee, C. G., & Huettel, S. A. (2012). The functional neuroanatomy of decision making: Prefrontal control of thought and action. Brain research, 1428, 3-12.

[38] Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neuroscience and biobehavioral reviews, 37, 1297-1310.

[39] Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience and biobehavioral reviews, 26, 321-352.

[40] Gazzaniga, M. S., Ivry, R. B., & Mangunm, G. R. (2004). Cognitive neuroscience: The biology of the mind. (2nd ed.). New Yourk: Norton

[41] Watanabe, K., Hikosaka, K., Sakagami, M., & Shirakawa, S. I. (2002). Coding and monitoring of motivational context in the primate prefrontal cortex. Journal of Neuroscience, 22(6), 2391- 2400.

[42] Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of Anterior Cingulate Cortex function. Neuron, 79, 217-240.

[43] Sutton, R. S., & Barto, A. C. (1998). Reinforcement learning: An introduction. Boston: MIT Press.

[44] Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain, 123, 2189-2202.