Neuroscience Letters 618 (2016) 110–114

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research paper

The role of time in conflict-triggered control: Extending the theory of response-conflict monitoring

Sareh Zendehrouh[a,*], Shahriar Gharibzadeh[b], Farzad Towhidkhah[b]

[a] School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5746, Tehran, Iran
[b] Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran

HIGHLIGHTS

- We present a computational model for a proposed cost-conflict monitoring system.
- We use the model to simulate human performance in a learning task.
- The model integrates reinforcement learning and conflict monitoring theories.
- The results indicate that the model can simulate human performance in the task.
- We show that the model can account for both the FRN and the ERN.

ARTICLE INFO

Article history:
Received 5 December 2015
Received in revised form 8 February 2016
Accepted 29 February 2016
Available online 2 March 2016

Keywords:
Cost
Reward
Goal-directed behavior
Reinforcement learning
Model-based RL

ABSTRACT

Flexible goal-directed behavior requires monitor-control networks to detect the need for behavioral adjustments and to implement the required regulations. Among event-related brain potentials related to the function of such networks is the feedback-related negativity (FRN), which is detected in trial-and-error learning tasks. Conflict monitoring theory (CMT) as one of the influential theories of such networks cannot describe the FRN. Recently, we have proposed a cost-conflict monitoring system that extends the CMT. The cost-conflict monitoring holds that the monitoring system can detect conflict signal, but the conflict is over the costs of alternative outcomes of the selected action rather than the response conflict as proposed by the CMT. In the cost-conflict monitoring, cost functions are computed based on waiting times from the response to feedback delivery and from these quantities a conflict signal is derived. Here, we present a computational realization of such cost-conflict monitor-controller network. We utilize this computational model to simulate existing human performance and ERP data of a trial-and-error learning task. The model successfully simulated the behavioral data and FRN signals under different conditions in this task.

1. Introduction

Monitoring and evaluating outcomes of actions and applying appropriate cognitive regulations and behavioral adaptations play a crucial role in decision-making and goal-directed behavior. In monitor-control loops responsible for such functions [1], the monitoring system detects the need for control, whose realization seems to be reflected in some components of event-related potentials (ERPs) [2]. The feedback-related negativity (FRN) as one of these components is observed in gambling and trial-and-error learning tasks [3] and peaks between 230 and 330 msec after feedback presentation [4].

The reinforcement learning (RL) account of such monitor-control loops [5,6] draws on the evidence indicating the resemblance between the phasic activity of dopamine neurons in the brain stem nuclei and reward prediction errors (RPEs) in temporal difference models of computational RL [7]. According to this theory, the output of the monitor located in the basal ganglia is conveyed to frontal areas in the form of phasic activity of dopamine neurons or RPE. The theory holds that the anterior cingulate cortex (ACC) uses these RPEs to improve performance and generates the FRN when receives negative RPEs, indicating the occurrence of an unexpected error feedback [6]. According to another competing theory, the conflict-monitoring theory (CMT) [8,9], the ACC moni-

* Corresponding author at: Sareh Zendehrouh, School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5746, Niavaran Sqr, Tehran, Iran.
E-mail addresses: sareh.zendehrouh@gmail.com, szendehrouh@ipm.ir (S. Zendehrouh).

http://dx.doi.org/10.1016/j.neulet.2016.02.062
0304-3940/© 2016 Elsevier Ireland Ltd. All rights reserved.

tors for conflicts during motor response generation as an index of the need for further control. The CMT suggests that the ACC acts as a response-conflict monitor that detects the simultaneous activation of competing motor responses, measures the conflict and sends this information to the dorsolateral prefrontal cortex to implement compensatory adjustments [8,9]. Nevertheless, the CMT cannot account for the FRN [6,9] which emerges when there is no activated motor response [1]. Although, the ACC was first discovered to be activated by the conflict between response representations, further investigations showed that the ACC can also be activated by conflicts between representations like semantic or conceptual [10,11]. It has been suggested that conflict might arise anywhere in the information processing system and might not be confined to the conflict between responses. In this view, the ACC detects a conflict between mutually active, competing representations and employs the dorsolateral prefrontal cortex to deal with the conflict [12]. Recently, we have proposed a cost-conflict monitor alongside the response-conflict monitor in the brain, which can extend the CMT to explain the FRN [13]. One of the ideas behind the proposed model of cost-conflict monitor is the ability of the brain in making retrospective reevaluations.

We specifically hypothesized that the subjective estimations of costs for likely outcomes of the selected action can be considered as the competing representations in the post response period. There is a consensus among neuroscientist that the passage of time can discount the value of a reward. Reward discounting can happen across several timescales [14]. Behavioral data from human beings and animals regarding reward discounting match well with a hyperbolic function of time [15]. In addition, investigation shows that the time passage also reduces the aversiveness of a negative event and obeys a hyperbolic function [16]. Therefore, because of the time interval between response selection and feedback presentation, each outcome carries a cost. Here we show that if humans revise their estimation of cost after receiving a feedback, the simultaneous activation of mutually exclusive estimations may lead to a cost-conflict signal.

2. Proposed model

2.1. The simulated task

In this paper, we have simulated the experimental data (Experiment 3) of the Holroyd et al. study [17] which is a trial-and-error learning task. Fifteen subjects had participated in the task. For every participant, there were three blocks of 300 trials. Thus, 900 trials were obtained for each participant. On each trial, a visual image of an object or an animal is displayed as a stimulus. Participants were asked to make a response by pressing a left mouse button (left response) or a right mouse button (right response). Response selection is then followed by one of six feedback conditions. Thus, learning occurs through trial-and-error efforts. For each stimulus, there is a probability of receiving a reward (on 20%, 50%, or 80% of the trials) only for the optimal response. The 80%, 50%, and 20% reward conditions are considered as the expected, control, and unexpected conditions, respectively [17].

2.2. The structure of the model

Reinforcement learning is one of the learning methods that has rich links with the neuroscience of decision-making and goal-directed behavior. With the sub-aim of conflating RL and conflict monitoring accounts of monitor-control loops, we used the principles of computational RL as the main basis for the proposed model. The original model of the RL theory uses temporal difference (TD) methods of the computational RL [6]. The TD mechanism is a model-free method that learns cached values of actions and better accounts for habitual behaviors or learning of stimulus-response (S-R) associations, while model-based methods learn the model of the environment to evaluate candidate actions and can better describe goal-directed behaviors [18,19] or learning of response-outcome (R-O) contingencies [20]. In a novel environment in which S-R mappings have not been learned [21] response selection is largely goal-directed and is controlled mainly by R-O associations [22]. Therefore, model-based RL is a better modeling tool for simulating trial-and-error learnings.

2.2.1. Response generation and value estimation

In the RL framework, a transition matrix denoted by$T(s, a, s')$ represents the probability of reaching state s' after taking action a at state s [23]. After each state transition that occurs after receiving a feedback, the state transition matrix is updated as follows:

$$T\left(s, a, \hat{s}\right) = \begin{cases} (1-\phi)T\left(s, a, \hat{s}\right) + \phi \ \ if\, \hat{s} = s' \\ (1-\phi)T\left(s, a, \hat{s}\right) \qquad otherwise \end{cases} \tag{1}$$

where, s is the stimulus state, a is the taken action in that state, and s' is the reaching state. $0 < \phi < 1$ is the update rate of the transition matrix.

The value of each state-action pair ($Q\,(s, a)$) is updated using the transition matrix and the received feedback:

$$Q(s, a) = \sum_{s'} T\left(s, a, s'\right)\left(r(s, a, s') + \gamma V(s')\right) \tag{2}$$

where r is the amount of the reward (=1 and 0 for a rewarding and non-rewarding feedback, respectively), γ is the discount factor, and $V(s')$ is the reaching state value (coded as +1 or −1 for rewarded (‘*re*’) and non-rewarded final state (‘*nr*’), respectively). The value of the stimulus state is assigned as follows:

$$V(s) = \max_a Q(s, a) \tag{3}$$

The action selection probabilities are calculated using the Softmax function [23]:

$$P_i = \frac{\exp(Q(s, a_i)\tau(s))}{\sum_{j=L,R}\exp(Q(s, a_j)\tau(s))} \tag{4}$$

where $Q(s, a_i)$ is the value of each state-action pair and τ is the inverse-temperature parameter in the Softmax function. The left and right response units are denoted by L and R, respectively.

The computed probabilities in Eq. (4) are used for response generation based on leaky competing accumulator [24] models of decision making in two-alternative-forced-choice tasks. The activation function of the response units obeys the following rule:

$$\Delta f_i^t = (E_i^t\theta^t + 0.04 - E_i^t f_i^t - I_i f_j^t)dt + N(0, .01) \tag{5}$$

where f_i^t and f_j^tshow the activity levels of a response unit and its competing response unit at time t, respectively.E_i^t and I_i show the excitatory and the inhibitory weights to response unit i, respectively. θ^t is a function that equals zero when the activity of either response units exceeds a decision threshold (response generation) and otherwise equals one. $dt(= 0.01)$ is a time constant, and $N\,(0, .01)$is a Gaussian noise with zero mean and standard deviation of .01.

2.2.2. Performance monitoring and control

The monitoring part in the model monitors the performance of other blocks and detects the need for further adjustments. Specifically, the monitoring mechanism detects the occurrence of unexpected outcomes and conflicts. When there is a need