Neural Networks 71 (2015) 112–123

Contents lists available at ScienceDirect

Neural Networks

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

A new computational account of cognitive control over reinforcement-based decision-making: Modeling of a probabilistic learning task

Sareh Zendehrouh

School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5746, Tehran, Iran

HIGHLIGHTS

- Modeling decision-making from the perspectives of dual-system and cognitive control.
- The model simulates human performance on a variant of probabilistic learning task.
- The model addresses existing theories about the ERN and FRN components of ERP.
- The results show that the ERN is best described by the RL-ERN theory.
- The FRN is best described by a hypothetical cost-conflict signal.

ARTICLE INFO

Article history:
Received 13 September 2014
Received in revised form 17 May 2015
Accepted 13 August 2015
Available online 20 August 2015

Keywords:
Cognitive control
Reinforcement learning
Goal-directed behavior
Dual system theory
Cost function
Probabilistic learning task

ABSTRACT

Recent work on decision-making field offers an account of dual-system theory for decision-making process. This theory holds that this process is conducted by two main controllers: a goal-directed system and a habitual system. In the reinforcement learning (RL) domain, the habitual behaviors are connected with model-free methods, in which appropriate actions are learned through trial-and-error experiences. However, goal-directed behaviors are associated with model-based methods of RL, in which actions are selected using a model of the environment. Studies on cognitive control also suggest that during processes like decision-making, some cortical and subcortical structures work in concert to monitor the consequences of decisions and to adjust control according to current task demands. Here a computational model is presented based on dual system theory and cognitive control perspective of decision-making. The proposed model is used to simulate human performance on a variant of probabilistic learning task. The basic proposal is that the brain implements a dual controller, while an accompanying monitoring system detects some kinds of conflict including a hypothetical cost-conflict one. The simulation results address existing theories about two event-related potentials, namely error related negativity (ERN) and feedback related negativity (FRN), and explore the best account of them. Based on the results, some testable predictions are also presented.

1. Introduction

The decision-making process, by which animals and humans select action among several alternatives, has been the subject of active research in many disciplines. Recent work in this field has given rise to the dual-system theory of decision making, which states that this process is conducted by two main systems: a deliberative "goal-directed" system controlled by response–outcome associations (R–O) and a relatively automatic or "habitual" one controlled by stimulus–response (S–R) mappings (Balleine & Dickinson, 1998; Daw, Niv, & Dayan, 2005; Valentin, Dickinson, & O'Doherty, 2007). The analysis of decision-making process from the computational perspective of reinforcement learning (RL) has given special attention in the past decade. In RL domain, the habitual behaviors are connected with model-free methods of RL, in which appropriate actions are learned through trial-and-error experiences. The idea behind the existence of a model-free system in the brain stems from the resemblance between phasic increases and decreases in firing rates of midbrain dopamine (DA) neurons and a reward prediction error (RPE) in model-free algorithms of RL (Suri, 2002). The model-free algorithms that are appropriate tools for modeling habit-driven stimulus–response

E-mail addresses: szendehrouh@ipm.ir, sareh.zendehrouh@gmail.com.

http://dx.doi.org/10.1016/j.neunet.2015.08.006

associations, cannot account for some goal-directed behavioral observations as devaluation sensitivity or latent learning (Bornstein & Daw, 2011). The current literature claims that goal-directed behaviors are associated with model-based methods of RL, in which actions are selected using a cognitive map or a model of the environment (Daw et al., 2005; Gläscher, Daw, Dayan, & O'Doherty, 2010; Simon & Daw, 2011). Work by some groups suggests that goal-directed and habitual behaviors are conducted by separate and parallel-operating model-free and model-based systems in the brain (Bornstein & Daw, 2011). However, others propose that these two systems may work together and are not completely separate (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Gutkin & Ahmed, 2012; Simon & Daw, 2011).

From the perspective of cognitive control, successful decision-making requires the organism to monitor the consequences of actions and decisions and to detect the failures in performance. Decision-making changes the state of the environment or the organism known as outcome. The differences between observed outcomes and expected ones are monitored by performance monitoring systems and are sent to brain areas responsible for control and optimization of current and/or future behaviors (Ullsperger, Danielmeier, & Jocham, 2014). Thus, a monitor-controller system is activated in such processes. The computational model of RL-ERN theory by Holroyd and Coles (Holroyd & Coles, 2002) is one of the influential models for such systems. This model links the impact of the observed outcome deviations from expected ones on neurons in the anterior cingulate cortex (ACC) to changes in the event-related brain potentials (ERPs), e.g. the ERN[1] that follow erroneous responses and the FRN[2] that follows positive/negative outcomes. The model of the RL-ERN as a model-free algorithm uses the dopaminergic RPE signals conveyed by mesolimbic pathways to optimize its performance and predictions. The RL-ERN theory holds that phasic decreases in dopamine activity disinhibit motor neurons in the ACC, thereby producing the ERN. In reinforcement learning (RL) terms, the ERN is generated when a negative RPE is sent to the ACC indicating the occurrence of an event that is worse than expected. Similarly, an FRN is elicited when an unexpected negative feedback occurs (Holroyd, Yeung, Coles, & Cohen, 2005). Another model, the predicted response–outcome (PRO) model by Alexander and Brown (Alexander & Brown, 2011) which is a generalization of RL algorithms, contains a model-based part that serves as an immediate outcome predictor to control the behavior, and a model-free part that learns a timed prediction of the expected time of outcome occurrence. The PRO model, which has some parts that operate like a model-based controller, holds that medial prefrontal cortex (mPFC) signals differences between timed outcome predictions and the actual outcomes. Although, these difference signals resemble RPEs in the RL-ERN theory, the PRO model suggests that such difference signals (surprise signals) are computed internally by the mPFC (Todd, Hills, & Robbins, 2012). Specially, the model holds that the mPFC activity represents the amount of negative surprise signal (unexpected non-occurrence of a predicted outcome). Therefore, the ERN and FRN are the manifestation of such signaling. However, such a description for ERN generation is inaccurate, because the actual outcome may not be determined at the time of response generation. This is in accordance with their research in Alexander and Brown (2011) which does not contain direct simulations of ERN component (Zendehrouh, Gharibzadeh, & Towhidkhah, 2013, 2014a). Moreover, this model is not essentially based on dopamine signals (Todd et al., 2012). Another prominent model, which does not belong to the RL category, is the model of conflict monitoring theory. This theory suggests the existence of a conflict monitor-controller system in the brain that monitors for the occurrence of conflict at the response level and uses this information to adjust the performance of the controller (Botvinick, Braver, Barch, Carter, & Cohen, 2001). In this model, simulated ERN is defined based on conflict signals that the ACC detects through monitoring the amount of energy in the motor cortex during action selection (Yeung, Cohen, & Botvinick, 2004). This theory in its original form cannot account for the FRN (Holroyd et al., 2005; Yeung et al., 2004). Recently, a hypothetical cost-conflict monitor has been proposed that extends this theory and can describe the FRN based on the conflict between expected costs of the selected action (Zendehrouh, Gharibzadeh, & Towhidkhah, 2014b).

In this paper, a new computational model for a variant of probabilistic learning task is given. While the proposed model, based on dual system theory of decision-making, simulates behavioral data from this experiment, it also explores the best description for ERN and FRN components that matches the empirical data. The simulation results show that the simulated ERN is better matched with the RPE signals in model-free part of the proposed model consistent with the RL-ERN description for this component. However, the amplitudes of the simulated FRN based on RPEs at the time of feedback onset are not so close to the empirical FRN amplitudes. Results also show that the FRN is better simulated by a hypothetical cost-conflict monitor-controller.

2. Materials and methods

2.1. Probabilistic learning task

The probabilistic learning task or the PLT is a trial-and-error learning task where an arbitrary visual image is presented to the participants on each trial. The participants press one of two buttons in response to that image and receive a feedback indicating receiving or missing a small amount of money. Here the data of Morris et al. (Morris, Heerey, Gold, & Holroyd, 2008; Morris, Holroyd, Mann-Wrobel, & Gold, 2011) are simulated, wherein each stimulus is probabilistically associated with the proper response on either 100%, 80%, or 50% of situations. Participants are not provided with the appropriate stimulus–response mappings. Instead, they have to deduce them by trial and error. Each block of the experiment consists of a new set of six stimuli. In summary, each stimulus and its accompanying outcome belongs to one of the following categories. (1) 100% mappings condition: The left button is the proper response for one of the six stimuli on 100% of trials within a block. The right button is the proper response for another stimulus in the same way. (2) 50% mappings condition: For two other stimuli, a random feedback was delivered. Therefore, the participants were rewarded on 50% of the trials and penalized on 50% of the trials. (3) 80% mappings condition: for one of the two remaining stimuli, a left button is the proper response on 80% of the trials (valid trials) and a right button is the proper response on 20% of the trials (invalid trials). For the other stimulus, a right button is the appropriate response on 80% of the trials (valid) and a left button is the appropriate on 20% of the trials (invalid) (Nieuwenhuis et al., 2002).

2.2. Proposed model

As mentioned earlier, both model-based and model-free methods are used in concert with each other to simulate human performance in the task. The structure of the model is depicted in Fig. 1.

[1] Error related negativity or error negativity (ERN/Ne) is an ERP component that begins near the time of the erroneous response and peaks about 100 ms later in speeded response time tasks (Gehring, Goss, Coles, Meyer, & Donchin, 1993).

[2] The feedback related negativity (FRN) is a negative-going component observed 230 to 330 msec following outcome presentation (Miltner, Braun, & Coles, 1997) in gambling and trial-and-error learning tasks (Holroyd, Hajcak, & Larsen, 2006).