Neuroscience Letters 685 (2018) 179–184

Contents lists available at ScienceDirect

Neuroscience Letters

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

Research article

Feedback delay impaired reinforcement learning: Principal components analysis of Reward Positivity

Hang Yin[a,b,1], Yu Wang[c,1], Xukai Zhang[a,b], Peng Li[a,d,*]

[a] Brain Function and Psychological Science Research Center, Shenzhen University, Shenzhen, China
[b] Research Center of Brain and Cognitive Neuroscience, Liaoning Normal University, Dalian, China
[c] School of Education and Psychology, Tianjin University of Sport, Tianjin, China
[d] Shenzhen Key Laboratory of Affective and Social Cognitive Science, Shenzhen University, Shenzhen, China

ARTICLE INFO

Keywords:
FRN
Reward Positivity
Feedback delay
PCA
Reinforcement learning

ABSTRACT

An immediate feedback after action facilitated reinforcement learning in dynamically varying environments. With several seconds delay, a series of event-related potential (ERP) studies have recently conducted to explore how delayed feedback influences learning processes and corresponding brain activities by measuring the Reward Positivity and N170 component. However, it remains unclear how does our brain process a feedback that is delayed longer and interrupted by other trials. In the present study, participants were asked to undertake a time-estimation task in two different conditions. Feedback was presented right after their actions in the immediate feedback condition, while it was presented after another five trials in the delayed feedback condition. By recording feedback related activities, we aim to test whether, or not, delayed feedback impairs reinforcement learning, the Reward Positivity and N170 amplitude. The behavioural results show that delayed feedback can reduce behavioural adjustment efficiency from trial-to-trial. To reduce component overlapping, we adopted the temporospatial principal components analysis (PCA) to separate the Reward Positivity from other ERP components. Results indicate that the Reward Positivity is decreased in the delayed feedback condition compared to the immediate feedback condition, however, no difference of N170 amplitude is found between the two conditions. These results indicate that delayed feedback impairs reinforcement learning process in terms of behavioural adjustment and brain activities even though these feedbacks are truly associated with participants' previous actions.

1. Introduction

Learning from the feedback of past performances is an essential ability of humans, it helps us to guide future behaviours to draw on the advantages and avoid disadvantages in the dynamic environment. A negative reward prediction error (RPE) signal occurs when the outcome feedback is worse than expectation and a positive RPE signal occurs when the feedback is better than expectation[1]. In cognitive neuroscience, the event-related potential (ERP) has been widely used to detect brain activities during feedback processing ([2]; for review, see René [3]). A great number of studies have consistently found that an ERP component, feedback-related negativity (FRN), is associated with the outcome evaluation and RPE processing [4,5]. FRN is a negative-going electrical potential that occurs 200 ms–350 ms after feedback stimuli onset is present and peaks at the frontocentral midline [6]. FRN amplitude was commonly measured by the difference wave between negative feedback and positive feedback, thus reflects the main effect of feedback valence [1]. According to the early reinforcement learning error-related negativity [4], the difference wave is mainly driven by negative RPE signals. However, recent evidence has indicated that the difference wave is more sensitive to positive RPE than negative RPE and was named Reward Positivity (RewP)[2] [7–10]. There is also evidence that the FRN component is sensitive to the absolute size of a RPE signal rather than negative or positive RPE in particular [11]. However, this statement was not supported by some following studies (*e.g.* [1,12]).

Feedback plays an essential role in reinforcement learning,

* Corresponding author at: No. 3688, Nanhai Road, Nanshan District, Shenzhen, 518060, China.
E-mail address: peng@szu.edu.cn (P. Li).

[1] Shared first author.

[2] For clarification, in the current study, the classical FRN effect was named "FRN" when it was measured on the original waveform and named "RewP" when it was measured by the difference wave between negative feedback and positive feedback. Additionally, the positive deflection within the classical FRN time window after PCA was named "PCA-FRN".

https://doi.org/10.1016/j.neulet.2018.08.039
Received 11 January 2018; Received in revised form 3 August 2018; Accepted 27 August 2018
Available online 28 August 2018
0304-3940/

however, it is common that a decision-maker cannot always obtain the outcome immediately following their actions. Instead, feedback is often delayed for seconds, days or longer. A number of studies have indicated that the processing of these feedback data with different waiting timings may engage different brain mechanisms. Specifically, the processing of the immediate feedback recruited the striatum [13,14], while the processing of the delayed feedback was supported by the medial temporal lobe (MTL), primarily the hippocampus [15]. Accordingly, several ERP studies have also focused on the time course of feedback processing with waiting timings manipulated between action and feedback. Weinberg et al. [16] adopted a forced choice gambling task to analyse the effect of feedback delay on the FRN. Their results suggest that the FRN amplitude difference is diminished at the long delays (6 s after response) when compared to the short delays (1 s after response). Moreover, Peterburs et al. [17] applied a probabilistic learning task to investigate the FRN effect modulated by increasing feedback delays, *i.e.*, short delay (500 ms), medium delay (3500 ms), or a long delay (6500 ms). They found a negative linear relationship between the amplitude of the difference waves between negative feedback and positive feedback and the feedback delay time. The authors proposed that the varied RewP effect might reflect the gradual brain activity shift from the striatum to hippocampus. Furthermore, Arbel et al. [18] suggested that the FRN and N170 are the two important ERP components able to capture neural activity in the striatum and MTL, respectively, when an individual learned from immediate feedback and delayed feedback. Yet one of the previous studies in our group found that P300, rather than the RewP, is sensitive to delayed time [19]. Thus, it remains arguable whether FRN (or RewP) is sensitive to the delay timing.

In the aforementioned studies, delayed feedback still provides learning information even though it is delayed for several seconds because each participant's working memory system is able to hold their action information until feedback presentation. In reality, however, it is highly possible that feedback is postponed longer than several seconds, therefore, a decision-maker hardly can hold previous action information in working memory, especially when the memory association between current action and corresponding feedback is interrupted by other tasks in hand. The primary goal of the present ERP study is to investigate brain activities for processing delayed feedback that provides rare learning information. In a time-estimation task, participants received feedback about their performance immediately or delayed after five more trials. This particular manipulation was adopted because of two points: First, in the delay condition, we provided participants the feedback, which was associated with response occurred five trials ago, right after their sixth response with the aim of keeping consistency between the delayed feedback condition and immediate feedback condition, *i.e.* every feedback stimulus was delivered after a key pressing in both conditions. Second, the manipulation of five trials delay was to simulate long feedback delay that decrease the memory association between response and feedback in real life. In the delayed feedback condition, participants had to remember their response from five trials ago and keep updating items in working memory if they tried to learn from feedback in a particular trial. Therefore, the delayed feedback condition was analogous to an N-back task that is widely used as a working memory measure [20]. When N = 3, the N-back test is quite difficult for participants [21], let alone when testing at a 5-back level as in the current paradigm.

In addition to the RewP component and N170, P300, a positive deflection that peaked at central-posterior electrodes on scalp was also frequently observed in feedback processing [22–24]. Although convergent evidence demonstrated that the RewP is linked to reinforcement learning process, several research groups argued that P300 rather than RewP could predict following behavioural adjustment after feedback ([23,25]; for review, see Luft [26]). As mentioned above, one of our previous studies have shown that several seconds delay modulated the feedback related P300 amplitude rather than the RewP [19]. Notably, the RewP component was found to be overlapped frequently by P300 in the literature [8]. Therefore, we adopted a temporospatial PCA to reduce contamination of other ERP components on the RewP [8,27].

In the present study, we firstly predicted that participants could not learn from delayed feedback behaviourally when it was interrupted by other trials and this would be observed in behavioural adjustment data. Moreover, given that the RewP component was associated with reinforcement learning [4,28], we hypothesised that this component would be reduced in delayed condition by more than in the immediate condition. Furthermore, recent studies have associated the N170 component with feedback processing in MTL [18]. It would be interesting to test whether, or not, a long delay, which hampered working memory, would impact the feedback-related N170 effect. No enhanced N170 amplitude in the delay feedback condition was expected to be observed because the present manipulation impaired the memory and learning in this condition.

2. Methods

2.1. Participants

Twenty healthy undergraduate students (nine females, age = 21.3 y ± 1.49 y) participated in our experiment. They are all with normal or corrected-to-normal vision and have no hearing loss or history of neurological disorders. Participants were informed to undertake a time estimation task, and their payment all depended on their performance, each subject received 30–40 Chinese Yuan (about 5–7 US dollars). This study was approved by the local ethical committee of Shenzhen University. All participants gave their written informed consent before the experiment.

2.2. Experimental procedure

The time estimation task applied in our experiment was modified based on the classic time estimation task [6]. Participants were instructed to estimate the duration of 1 s by pressing the Space button in the keyboard. As shown in Fig. 1a, at the beginning of each trial, a fixation (+) appeared in the centre of the screen (500 ms). After a blank screen (600–800 ms), an auditory stimulus (1500 Hz, 50 dB, lasting 50 ms) was released to participants by earphones. Participants were told to press the button as soon as when they believed that 1 s had passed after the auditory stimulus's presence. Feedback stimulus was then presented after a random duration (600–1000 ms) following their response. Finally, if participants' reaction time (RT) fell within a time window that centred around 1 s, they received "√" mark (1000 ms) as the positive feedback, indicated that the response was correct and won a reward of ¥0.5. Otherwise, a "×" sign indicated their incorrect response and no reward as the negative feedback. Notably, the initial correct time window was set as 900–1100 ms and the window size was adjusted by the performance of the participants (*c.f.*, [6]). In the immediate condition, the width of the window decreased by 10 ms in the next trial if their response was correct and the width of the window increased by 10 ms if their response was incorrect. In the delay condition, however, it is more difficult for participants to receive 50% correct feedback because the delay between the feedback and the relevant response made learning more difficult. Thus, the width of the window decreased by 6 ms in the next trial if their response was correct and the width of the window increased by 18 ms if their response was incorrect. Based on these manipulations, the probability of positive feedback was kept at about 50% in both of immediate and delay condition. Before the formal experiment, participants conducted 20 practice trails. The inter-trial interval was set randomly from 1000 ms to 1500 ms by presenting a blank screen.

The whole experiment was divided into two blocks and each type of condition was manipulated separately in each block. In the immediate feedback block, participants received feedback of their accuracy after 600–1000 ms following the response. In contrast, in the delayed