International Journal of Psychophysiology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

Disappearance of self-serving bias: Reward positivity reflects performance monitoring modulated by responsibility attribution in a two-person cooperative task

Peng Li[a,b], Hang Yin[a], Huyan Xu[c], Yi Lei[a,*], Hong Li[a,b,d]

[a] Brain Function and Psychological Science Research Center, Shenzhen University, Shenzhen, China
[b] Shenzhen Key Laboratory of Affective and Social Cognitive Science, Shenzhen University, Shenzhen, China
[c] School of Education Science, Anhui Normal University, Wuhu, China
[d] Center for Language and Brain, Shenzhen Institute of Neuroscience, Shenzhen, China

ARTICLE INFO

Keywords:
Reward positivity
Self-serving bias
Responsibility attribution
Performance monitoring

ABSTRACT

Performance monitoring plays a virtual role in individual reinforcement learning. However, it remains unclear how responsibility attribution modulates the individual monitoring process in a social cooperative context. In the present study, 46 participants received feedback on the team's monetary outcome, teammate performance, and their own performance sequentially for a two-person task. Using event-related potential (ERP), we analyze brain activity in response to performance monitoring during team and self feedback, indexed according to reward positivity (RewP). Overall, the participants reported a modest tendency towards causal attribution in terms of taking more responsibility for negative rather than positive team-feedback, thus indicating an opposite pattern to the so-called *self-serving bias* phenomenon. Based on post-experiment responsibility attribution, participants were further divided into a 'Modest' group ($N = 23$) who reported more responsibility for team failure than success, and an 'Ordinary' group ($N = 23$) who made comparable attribution irrespective of team outcome. The ERP results show that there is no difference in RewP amplitudes between the two groups when the participants were processing the team's monetary feedback. However, the observed RewP amplitudes are notably different in the Modest group when processing self-performance feedback at different levels of responsibility attribution. These findings demonstrate that neural activity during performance monitoring does not differ between the two groups. However, using different responsibility attribution tendencies does affect brain activity during individual performance monitoring. The observed RewP effect sheds light on the automatic and implicit evaluation of one's own performance in a social cooperative context.

1. Introduction

When humans behave in a goal-directed manner, a flexible performance monitoring system serves to select advantageous actions and compensate for errors made with a minimal amount of effort (Ullsperger et al., 2014). Compared with that in a non-social context, performance monitoring in an interpersonal influence context requires considering not only one's own outcome produced by his/her performance but also the influence on the outcome of others in the team (Rilling and Sanfey, 2011). Cooperating with others is undoubtedly beneficial to survival for animals and humans in complicated environments. According to social norms, an individual is assumed to make a contribution to social teams in a responsible way, rather than merely consuming public good as a 'free-rider' (Buckholtz and Marois, 2012). Thus, it is crucial for individuals to judge the social responsibility aspect of their performance on the basis of social prescriptions (Schlenker et al., 1994). Moreover, people are also motivated to make causal attribution of their individual contributions to team success/failure so as to enhance or protect self-esteem (Weiner, 1985; Zuckerman, 1979). A perceived sense of responsibility before performance or responsibility attribution after performance can, in return, dramatically affect one's self-conscious emotions and the regulation of future behavior and decisions (Weiner, 1985; Leary, 2007; Schlenker et al., 1994; Tracy and Robins, 2004).

Causal attribution theory posits that locus, stability, and controllability are the three main properties of attribution in achievement-related tasks (Weiner, 1985). Image that two 'players' are performing a cooperative task that requires a sustained cognitive effort. Each player

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

https://doi.org/10.1016/j.ijpsycho.2018.09.002
Received 26 April 2018; Received in revised form 15 July 2018; Accepted 8 September 2018

could attribute the final outcome to their own performance, his/her teammate's performance, or external factors based on the locus of the causality. Typically, people make asymmetric responsibility attributions depending on the goodness of the team outcome. For instance, people tend to overstate their own contribution to team achievement and believe that they are less responsible for team failure than others. This is known as *self-serving bias* (Miller and Ross, 1975; Zuckerman, 1979; Forsyth, 2008; Duval and Silvia, 2002). Self-serving bias can be modulated by a range of factors, e.g. interpersonal relationships, behavioral privacy, culture differences, and personality (Shepperd et al., 2008; Greenberg et al., 1982; Arkin et al., 1980; Sedikides et al., 1998). For example, members with distant and not close relationships manifest self-serving bias by claiming more credit for team success and denying their responsibility for team failure (Sedikides et al., 1998; Campbell et al., 2000). When participants believe that their attributions will be revealed to other members in the group, they show less self-serving bias (Miller and Schlenker, 1985; Kudo and Numazaki, 2003). Moreover, cross-cultural research has found that people from Eastern cultures are less likely to exhibit self-serving bias than those from Western cultures (Heine and Hamamura, 2007; Yan and Gaier, 1994). In particular, people are likely to display self-enhancement with respect to traits and behavior that are emphasized within their own culture, e.g. 'independent' and 'unique' in Western cultures and 'agreeable' and 'cooperative' in Eastern cultures (Sedikides et al., 2003). However, the finding of differences between Western and Eastern cultures remains arguable. For instance, a review study has shown that samples from China and Korea were different to those in other sub-categories in East Asia and reveal similar self-serving bias when compared with North Americans (Mezulis et al., 2004). Later, Brown et al. (2009) also found that Chinese participants made comparable self-serving attributions as Americans. These somewhat inconsistent findings call for further studies on self-serving bias in Chinese samples.

Numerous studies have been conducted on responsibility attribution and self-serving bias. However, few of them focus on the relationship between responsibility attribution and individual performance monitoring in social cooperative contexts. It is not new that individuals evaluate their own performance (via internal monitoring or external feedback) in a trial-and-error task with the aim of maximizing profit using the reinforcement learning algorithm (Holroyd and Coles, 2002; Sutton and Barto, 1998). Previous event-related potential (ERP) studies have consistently associated the reward positivity (RewP) component (a difference waveform reflecting different brain activities in response to positive and negative feedbacks measured within 240–340 ms after the onset of feedback at frontal-central electrodes) with performance monitoring (Holroyd and Coles, 2002, for reviews, see Proudfit, 2015; Holroyd and Umemoto, 2016; Sambrook and Goslin, 2015). According to the influential reinforcement learning theory, RewP manifests the reward prediction error (RPE) signals generated by the dopamine system and conveyed to the anterior cingulate cortex (Holroyd and Coles, 2002; Holroyd and Umemoto, 2016). The RPE signals are indicative of the difference between the expected and received outcomes. It plays a key role in many types of reinforcement learning processes, including evaluating monetary outcome and performance monitoring (Holroyd and Coles, 2002; Li et al., 2010). Whether RewP reflects signed or unsigned RPE signals is still under debate. However, it has been repeatedly reported in the literature that the RewP amplitude is enhanced by a larger RPE (Arbel et al., 2013; Cohen and Ranganath, 2007; Chase et al., 2011; Luft et al., 2014).

To date, a number of ERP studies have investigated how a sense of responsibility before or during performance modulates brain activity when evaluating monetary outcome in a social context (Beyer et al., 2017; Li et al., 2010; Loehr et al., 2015). For instance, our previous ERP study (involving a dice-tossing task) showed that the feedback-related RewP amplitude is reduced when one works together with two teammates compared to when one performs the task alone (Li et al., 2010). Furthermore, the amplitude of the RewP reduction is found to be associated with the subjective rating of responsibility reported after the experiment. These findings suggest that diffusion of responsibility influences neural activity during outcome evaluation (also see Loehr et al., 2015). Beyer et al. (2017) further demonstrated that diffusion of responsibility can moderate the sense of agency and reduce RewP amplitude, rather than just displaying a post-hoc self-serving bias. As they manipulate sense of agency in different ways, the abovementioned studies collectively compare different brain responses to feedback under individual and cooperative conditions. However, to the best of our knowledge, no ERP study has yet examined how post-hoc responsibility attribution modulates individual performance monitoring in a cooperative context (with the sense of agency remaining unchanged before performance between conditions).

In the current ERP study, we aim to study if participants with different attribution tendencies have different brain response patterns to feedback and different behavioral adjustment strategies. For this purpose, we adapted a two-person cooperative task used in our previous functional magnetic resonance imaging study (Li et al., 2013). In the present study, participants were sequentially presented with team outcome, feedback on the teammate's performance, and their own performance feedback. Participants were asked to make specific responsibility attributions in each trial and an overall attribution assessment after the whole experiment had finished. By recording ERP signals, this modified paradigm allows us to separately determinate the participants' brain activities in the three feedback phases. Moreover, participants were split into two groups based on their post-hoc responsibility attribution tendencies to create a 'Modest' group, who reported more responsibility for team failure than success, and an 'Ordinary' group, who made comparable attributions for team failure and success. Given that people in collective cultures commonly reveal less self-serving bias than people in other cultures (Heine and Hamamura, 2007; Sedikides et al., 2003), we firstly hypothesized that, overall, the participants in our study tended to make non-self-serving responsibility attribution based on how much they contributed to team outcome. Further, we predicted that the Modest group would be more sensitive to monitoring self performance at different responsibility levels compared to the Ordinary group (as indicated by the RewP component). Additionally, we forecasted that the participants who exhibit modest attribution tendency would be more sensitive to their own error feedback, even when their team won, and would tend to make more behavioral adjustments in subsequent trials. This is in accordance with the previous finding that failure is likely to be attributed internally when people can improve their performance (Duval and Silvia, 2002).

2. Methods

2.1. Participants

Forty-six right-handed volunteers participated in this study (23 were females and the age range was 17–26 years with an average of 20.6 ± 2 years (mean ± standard deviation)). All participants had normal or corrected-to-normal vision, and reported no history of neurological or psychiatric illness. This study was approved by the local ethical committee in Shenzhen University. All participants gave their written informed consent prior to the experiment and received 70 Chinese Yuan (about 10 US dollars) for their participation.

2.2. Task

The whole experiment was composed of two sessions: (i) a behavioral test session, followed three days later by (ii) an EEG session. In the first session, participants were told to perform a classical 1-second time-estimation task on a computer. They received correct feedback on the screen if their estimate was within the allowed time-window (900–1100 ms), and incorrect feedback if they were outside the time-window. Actually, unbeknown to the participants, the time-estimation