Biological Psychology 149 (2020) 107775

Contents lists available at ScienceDirect

Biological Psychology

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

Dissociating the contributions of reward-prediction errors to trial-level adaptation and long-term learning

K.R. Lohse[a,b,*], M.W. Miller[c,d], M. Daou[c,e], W. Valerius[c], M. Jones[f]

[a] University of Utah, Department of Health, Kinesiology, and Recreation, United States
[b] University of Utah, Department of Physical Therapy and Athletic Training, United States
[c] Auburn University, School of Kinesiology, United States
[d] Auburn University, Center for Neuroscience, United States
[e] Coastal Carolina University, Department of Kinesiology, United States
[f] University of Colorado, Department of Psychology and Neuroscience, United States

ARTICLE INFO

Keywords:
EEG
Reinforcement learning
RewP
Adaptation

ABSTRACT

Reward positivity (RewP) is an EEG component reflecting reward-prediction errors. Using multilevel models, we measured single-trial RewP amplitude from trial-to-trial, while reward and prediction varied during learning. Sixty participants completed a category-learning task in either engaging or sterile conditions with the RewP time-locked to feedback. Sequential analysis of single-trial RewP showed its relationship to current and previous accuracy, and the probability of changing one's response to subsequent stimuli. Simulations show these effects can be explained in detail by the dynamics of participants' expectations according to principles of reinforcement learning. The single-trial RewP findings were consistent with previous literature linking RewP to reward-prediction error under reinforcement-learning theory. In contrast, the aggregate RewP was unrelated to the engagement manipulation or to delayed retention performance. Thus the present results provide a detailed computational account how RewP relates to acute adaptation, but suggest RewP plays little role in long-term learning.

1. Introduction

Reinforcement learning theory posits that individuals adjust their behavior in order to obtain rewards and avoid punishments (Sutton & Barto, 1998; Thorndike, 1927). These behavioral adjustments are driven by reward-prediction errors—the degree to which an actual reward differs from the learner's expectation (Schultz, 2017). Reward-prediction errors can result from rewards being better (positive) or worse (negative) than predicted, and both positive and negative prediction errors influence future behavior. Positive reward-prediction errors act as a signal within the brain to 'stamp in' a behavior, making the preceding action more likely to be selected in the future for a given state of the system (Holroyd & Krigolson, 2007; Palidis, Cashaback, & Gribble, 2018). Conversely, negative reward-prediction errors act as a signal within the brain to 'stamp out' a behavior, so that it can be avoided in the future.

In many studies of human learning, researchers measure reward-prediction errors through the proxy measure of the reward positivity (RewP) component derived from event-related potential (ERP) waveforms in electro-encephalograms (EEGs; Proudfit, 2015; Sambrook & Goslin, 2015). (This component has gone by other names including the feedback negativity, feedback-related negativity, feedback error-related negativity, and feedback correct-related positivity, but it is operationally and conceptually the same component as the RewP [Sambrook & Goslin].) Operationally, the RewP is a positive deflection in the ERP waveform that occurs 250 to 350 ms following positive feedback relative to negative feedback. The RewP exhibits a fronto-central scalp distribution, likely generated by anterior cingulate cortex. Conceptually, the RewP is believed to reflect a reward-prediction error (Holroyd & Coles, 2002). The positive prediction errors are likely transmitted to anterior cingulate cortex by a phasic increase in dopamine from the midbrain (Krigolson, 2018). In current practice, most researchers operationalize the RewP as a difference wave, specifically by averaging together ERPs from all trials with positive feedback, all trials with negative feedback, and then subtracting the latter average from the former. Alternatively, some researchers have looked at the RewP on a trial-by-trial level, operationalizing it as the voltage following feedback (Collins & Frank, 2018; Frömer, Stürmer, & Sommer,

[*] Corresponding Author at: 250 S 1850 E, Rm 258, Salt Lake City, UT, 84112, United States.
E-mail address: rehabinformatics@gmail.com (K.R. Lohse).

https://doi.org/10.1016/j.biopsycho.2019.107775
Received 22 January 2019; Received in revised form 23 September 2019; Accepted 23 September 2019
Available online 26 September 2019

2016). We refer to these two different measures as the aggregate RewP and single-trial RewP, respectively.

Converging evidence implicates both the aggregate and single-trial RewP as indices of reward-prediction error (Sambrook & Goslin, 2015). The difference-wave approach reliably yields a deflection in the waveform, indicating a strong sensitivity to the valence of feedback. The aggregate RewP has also been shown to be greater in response to larger rewards (e.g., \$0.5 vs. \$5.0). Thus, RewP is affected by both the sign and the magnitude of reward. Aggregate RewP is also sensitive to the participant's predictions (i.e., expectation of reward), being greater for unexpected outcomes than for expected outcomes (Holroyd & Krigolson, 2007; Sambrook & Goslin, 2015). The combination of positive dependence on reward and negative dependence on predictions implicates the single-trial RewP as a neural correlate of their difference, that is, reward-prediction error.

Currently, there is a limited understanding of how RewP relates to behavior over different timescales. Although a number of recent electrophysiological studies have investigated trial-by-trial dynamics of reward-prediction error (Chase, Swainson, Durham, Benham, & Cools, 2011; Collins & Frank, 2018; Fischer & Ullsperger, 2013; Frömer et al., 2016; Pedroni, Langer, Koenig, Allemand, & Jancke, 2011; Philiastides, Biele, Vavatzanidis, Kazzer, & Heekeren, 2010; Sambrook & Goslin, 2014, 2016; Sambrook, Hardwick, Wills, & Goslin, 2018), most work has focused on block- or session-level manipulations (for representative examples, see Bellebaum & Daum, 2008; Reinhart & Woodman, 2014; van der Helden, Boksem, & Blom, 2009). Moreover, there has been little to no research on how RewP relates to delayed retention or transfer, as opposed to immediate performance. The fact that much of the extant literature does not consider trial-by-trial dynamics or delayed retention and transfer tests is problematic for at least two reasons. First, analysis of trial-level (i.e., sequential) effects can be highly informative regarding the details of reinforcement learning mechanisms (Jones, Love, & Maddox, 2006; Jones, Curran, Mozer, & Wilder, 2013; Philiastides et al., 2010). Second, the relationship between the acute adaptation mechanisms (as postulated by reinforcement learning theory) and long-term learning is poorly understood. From the perspective of reinforcement learning theory, positive reward-prediction errors during practice drive adaption toward better performance. Reinforcement learning explains long-term learning as the accumulation of these adaptations (e.g., Sutton & Barto, 1998). These adaptations are also assumed to underlie generalization to novel stimuli, via overlap in stimulus representations (e.g., Jones et al., 2006). Thus, a straightforward prediction is that achieving a high level of performance during practice should be associated with better performance on subsequent retention and transfer tests. However, behavioral studies have shown that these two measures of performance are often uncorrelated or even negatively correlated (Kantak & Winstein, 2012; Pashler & Baylis, 1991).

The present experiment investigated the neurophysiological correlates of adaptation and learning in a perceptual categorization task. The primary aim was to evaluate the separate impacts of reward and prediction on the aggregate RewP and the single-trial RewP. We experimentally manipulated the value of reward between participants using a motivational game manipulation designed to enhance reward processing (Lohse, Boyd, & Hodges, 2016). To do this, we adapted a category learning task using complex visual stimuli called greebles (Gauthier & Tarr, 1997). In the "sterile" group, described in detail below, participants had to learn which of several responses corresponded to each family of greebles through trial-and-error categorization. In the "game" group, participants had to complete the same task, but rather than as a cognitive psychology experiment the task was framed as a game, "Goblin Quest", in which participants had to learn which of several "weapons" (responses) corresponded to each "goblin" (family of greebles). Building on previous work indicating that motivation enhances learning (Wulf & Lewthwaite, 2016), we hypothesized that increased motivation from the game manipulation would magnify representations of reward, yielding stronger reward-prediction errors (and thus magnified RewP).

In addition to the empirical study, we present and test a detailed analysis of computational reinforcement-learning models for trial-level dynamics of reward-prediction errors. Previous research on RewP has primarily used random feedback or stairstep procedures to hold reward probability to predetermined values (e.g., Holroyd & Krigolson, 2007). However, it is important to recognize that the probability of reward is not constant in natural learning environments. As learners become more skillful and knowledgeable in a task, success is more likely, and the accuracy of their predictions also increases. To the extent that RewP indexes prediction error, as opposed to just reward, it should be impacted by changes in learners' expectancy of success. Indeed, studies that have investigated trial-by-trial dynamics when learners' predictions are allowed to evolve freely over the course of learning reveal intriguing temporal dynamics of the single-trial RewP (Collins & Frank, 2018; Fischer & Ullsperger, 2013; Frömer et al., 2016; Philiastides et al., 2010; Sambrook & Goslin, 2016; Sambrook et al., 2018). For instance, Frömer et al. (2016) showed how the single-trial RewP displays characteristics of both a reward measure, being greater following accurate performances, and an expectancy measure, getting lesser as cumulative accuracy gets better. In the present study, we add to this research and extend it by (A) using mixed-effect regression analyses to show the relationship between the single-trial RewP and both preceding and subsequent behavior, (B) using simulations to show how the dynamics of the RewP can be explained in fine detail by a reinforcement learning model, and (C) including delayed retention and transfer tests to investigate whether observed short-term learning dynamics carry over to explain long-term learning.

1.1. Separating dynamics of prediction and reward in RewP

Because reward-prediction error is a difference between actual and expected reward, both factors are relevant to the interplay between prediction errors and learning. To the extent that RewP is sensitive to variation on the *reward* side, the most straightforward prediction is that single-trial RewP is more positive on correct trials than on incorrect trials. The corresponding property of the aggregate RewP, that the difference wave is positive, is well established and is the original basis for its interpretation as a correlate of reward-prediction error (Holroyd & Coles, 2002). However, reward-side effects on RewP also lead to several more detailed predictions. First, stronger subjective reward should produce larger reward-prediction errors, leading to greater behavioral adaptation (e.g., Cashaback, McGregor, Mohatarem, & Gribble, 2017). This is the assumption underlying the game manipulation: the gamified task would increase subjective reward, magnifying the RewP, and thus speeding learning. Similarly, to the extent that there are individual differences in strength of subjective reward, this assumption also predicts a positive correlation between aggregate RewP and training performance across participants (Grand, Daou, Lohse, & Miller, 2017; Holroyd & Krigolson, 2007). Moreover, if reinforcement learning mechanisms are important for long-term learning, and not just short-term adaptation, then the same correlation should be obtained for post-test performance (e.g., Abe et al., 2011). Finally, there is the parallel prediction at the within-participant level: If subjective reward varies across trials (separately for positive and negative feedback), there should be a positive correlation between the single-trial RewP on the current trial and the probability of repeating the current response the next time a stimulus from the same category is presented. This is because a greater RewP indicates the subjective value of that response given the stimulus will be more greatly increased (positive feedback) or more weakly diminished (negative feedback).

To the extent that RewP is sensitive to variation on the *prediction* side, three somewhat counterintuitive predictions derive directly from reinforcement learning theory. First, participants who have learned the task better will have greater expectation of reward when they choose a correct response and lesser expectation when they choose an incorrect