Biological Psychology 129 (2017) 265–272

Contents lists available at ScienceDirect

Biological Psychology

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

When theory and biology differ: The relationship between reward prediction errors and expectancy

Chad C. Williams[a,*], Cameron D. Hassall[a], Robert Trska[a], Clay B. Holroyd[b], Olave E. Krigolson[a]

[a] Centre for Biomedical Research, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada
[b] Department of Psychology, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada

ARTICLE INFO

Keywords:
Reinforcement learning
Reward positivity
Feedback-related negativity
Rescorla-Wagner learning rule
Prediction error
Dopamine

ABSTRACT

Comparisons between expectations and outcomes are critical for learning. Termed prediction errors, the violations of expectancy that occur when outcomes differ from expectations are used to modify value and shape behaviour. In the present study, we examined how a wide range of expectancy violations impacted neural signals associated with feedback processing. Participants performed a time estimation task in which they had to guess the duration of one second while their electroencephalogram was recorded. In a key manipulation, we varied task difficulty across the experiment to create a range of different feedback expectancies − reward feedback was either very expected, expected, 50/50, unexpected, or very unexpected. As predicted, the amplitude of the reward positivity, a component of the human event-related brain potential associated with feedback processing, scaled inversely with expectancy (e.g., unexpected feedback yielded a larger reward positivity than expected feedback). Interestingly, the scaling of the reward positivity to outcome expectancy was not linear as would be predicted by some theoretical models. Specifically, we found that the amplitude of the reward positivity was about equivalent for very expected and expected feedback, and for very unexpected and unexpected feedback. As such, our results demonstrate a sigmoidal relationship between reward expectancy and the amplitude of the reward positivity, with interesting implications for theories of reinforcement learning.

1. Introduction

Reinforcement learning in humans and other animals depends on the computation of prediction errors − discrepancies between the expected and the actual value of outcomes. Computationally, prediction errors are used to update the values of choice options so that over time behaviour is optimized to achieve the system's primary goal of maximizing reward (Rescorla & Wagner, 1972; Sutton & Barto, 1998; c.f. utilitarianism, Mill, 1863). Past findings with monkeys suggest that learning systems within the simian brain utilize neural prediction errors to optimize behaviour, with the primary supportive evidence being the scaling of the firing rate of the midbrain dopamine system in these animals in a manner predicted by reinforcement learning theory (Schultz, Dayan, & Montague, 1997; see also Amiez, Joseph, & Procyk, 2005; Matsumoto, Suzuki, & Tanaka, 2003; Matsumoto, Matsumoto, Abe, & Tanaka, 2007; Schultz, Tremblay, & Hollerman, 1998; Shidara & Richmond, 2002). For example, in a seminal study, Schultz et al. (1997) demonstrated that the firing rates of neurons within the midbrain dopamine system in monkeys mirrored the theoretical predictions of reinforcement learning: with learning, the dopamine neuron firing rates concomitantly decreased to rewards and increased to cues predicting the rewards. In humans, studies using both functional magnetic resonance imaging (Bray & O'Doherty, 2007; Brown & Braver, 2005; Haruno & Kawato, 2006; Jessup, Busemeyer, & Brown, 2010; Nieuwenhuis et al., 2005; Niv, Edlund, Dayan, & O'Doherty, 2012; O'Doherty et al., 2004; Roy et al., 2014; Silvetti, Seurinck, & Verguts, 2013; Tanaka et al., 2004; Tobler, O'Doherty, Dolan, & Schultz, 2006) and electroencephalography (Cohen & Ranganath, 2007; Eppinger, Kray, Mock, & Mecklinger, 2008; Ferdinand, Mecklinger, Kray, & Gehring, 2012; Hajcak, Moser, Holroyd & Simons, 2007; Hassall, MacLean, & Krigolson, 2014; Hewig et al., 2007; Holroyd & Krigolson, 2007; Holroyd & Coles, 2002; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Holroyd, Krigolson, Baker, Lee, & Gibson, 2009; Krigolson & Holroyd, 2007; Krigolson et al., 2011; Krigolson, Hassall, & Handy, 2014; Morris, Heerey, Gold, & Holroyd, 2008; Nieuwenhuis et al., 2002; Walsh & Anderson, 2012) have shown learning-related changes in the evoked responses to reward feedback that suggest that the underlying neural systems generating these signals are computing prediction errors. Specifically, the aforementioned studies in humans (and in monkeys) have shown a sensitivity of reward

* Corresponding author at: School of Exercise Science, Physical & Health Education, University of Victoria, P.O. Box 17000 STN CSC, Victoria, British Columbia, V8W 2Y2, Canada.
E-mail address: ccwillia@uvic.ca (C.C. Williams).

http://dx.doi.org/10.1016/j.biopsycho.2017.09.007
Received 14 March 2017; Received in revised form 6 June 2017; Accepted 14 September 2017
Available online 18 September 2017
0301-0511/

signals to expectancy – the difference between unexpected rewards and punishments elicit a larger neural response than the difference between expected rewards and punishments (e.g., Holroyd & Krigolson, 2007; Sambrook & Goslin, 2015).

Reward prediction error theories derive from early mathematical formalisms of reinforcement learning. Rescorla and Wagner (1972) proposed that surprising events should have more impact on behaviour than unsurprising events. They offered that the value of a given cue was the prediction, or the expectancy, of a subsequent outcome; as such, they defined a prediction error as the difference between the value of an outcome and the value of the cue that predicted the outcome. In mathematical models, for example, if a cue would lead with 100% confidence to a reward, its value would be 1, yet if the agent was unsure whether the cue would result in a reward (50% chance of reward), then the value would be 0.5. This position holds that larger differences between expected and outcome values lead to larger prediction errors. Rescorla and Wagner (1972) also proposed that the degree of learning is proportional to the magnitude of prediction errors, with larger and smaller prediction errors resulting in larger and smaller changes in value and behavior, respectively. On this account, the degree of learning from an outcome is linearly related to the expectedness of an outcome. Additionally, modern developments of the Rescorla-Wagner learning rule (e.g., temporal difference learning; Sutton & Barto, 1990; Sutton & Barto, 1998), continue to describe the relationship between learning and outcome expectedness to be linear. This prediction has received substantial empirical support. For instance, studies have shown that the magnitude of neural prediction error signals impacts the magnitude of behavioural adaptations on future trials within a re-occurring environment in that the larger the prediction error signal, the larger the behavioural adaptation (Cavanagh, Frank, Klein, & Allen, 2010; Cohen & Ranganath, 2007; Frank, Woroch, & Curran, 2005; Gehring, Goss, Coles, Meyer, & Donchin, 1993; Holroyd & Krigolson, 2007; Holroyd et al., 2009; Morris et al., 2008; Wessel, Danielmeier, Morton, & Ullsperger, 2012).

In principle then, neural systems for reinforcement learning should be sensitive to differing levels of expectancy deviation (i.e., differing prediction error magnitudes). Supporting this, Holroyd and Krigolson (2007) demonstrated that the amplitude of the reward positivity (formerly the feedback-related negativity), a medial-frontal component of the human event-related brain potential (ERP) involved in reward evaluation, scaled to outcome expectancy during performance of a time estimation task in which on each trial participants guessed the duration of one second and received feedback on their performance. They showed that the amplitude of the reward positivity for unexpected outcomes was larger than the reward positivity for expected outcomes. Importantly, they demonstrated that changes in response times were larger following incorrect trials than correct trials, as well as unexpected trials than expected trials, demonstrating that behavioural adaptations were related to the amplitude of the reward positivity. In a follow-up study that confirmed and extended this result, Holroyd et al. (2009) demonstrated that the reward positivity scaled across three levels of expectancy – expected (80%), control (50%), and unexpected (20%: see also Cohen, Elger, & Ranganath, 2007; Eppinger et al., 2008; Ferdinand et al., 2012; Hajcak et al., 2007; Hewig et al., 2007; Holroyd & Coles, 2002; Holroyd, Pakzad-Vaezi, & Krigolson, 2008; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Holroyd, Krigolson, & Lee, 2011; Kreussel et al., 2012; Liao, Gramann, Feng, Deák, & Li, 2011; Martin & Potts, 2011; Nieuwenhuis et al., 2002; Ohira et al., 2010; Pfabigan, Alexopoulos, Bauer, & Sailer, 2011; Potts, Martin, Burton, & Montague, 2006; Walsh & Anderson, 2011).

In contrast to these computational theories, biological processes are often non-linear. For example, non-linearity has been found in the endocrine system (Baldi & Bucherelli, 2005), in synaptic plasticity (Kerr, Huggett, & Abraham, 1994), and in neural communication (Foster, Kreitzer, & Regehr, 2002). Indeed, even midbrain dopamine signaling has been characterized as non-linear when manipulating reward expectancy (Fiorillo, Tobler, & Schultz, 2003) and reward magnitude (Schultz, 2016; Schultz et al., 2015; Stauffer, Lak, Kobayashi, & Schultz, 2016; Stauffer, Lak, & Schultz, 2014). For example, Stauffer et al. (2014) gave monkeys unpredictable rewards of varying magnitude (0.1–1.2 ml of juice). The authors asserted that, because the rewards could not be predicted, reward predictions were constant and near zero. Thus, they claimed, prediction error magnitudes were proportional to reward magnitudes. Interestingly, they observed that dopamine activation comported to a sigmoid-shaped utility function, such that extreme gains and losses resulted in relatively smaller changes in subjective value (see Bernoulli, 1738 /1954; Mas-Colell, Whinston, & Green, 1995).

Thus a relationship between reward expectancy and prediction error amplitudes is apparent, yet the issue of linearity has never been examined. In the present study, we investigated the relationship between reward expectancy and a neural correlate of reward evaluation, the reward positivity, across a range of expectancies from very expected to very unexpected. The reward positivity reflects the evaluation of reward feedback within the human medial-frontal cortex and is quantified as the difference between the 'positive' feedback waveform and the 'negative' feedback waveform (positive – negative; see Proudfit, 2015 for a review). Similar to Holroyd and Krigolson (2007), we employed a time estimation task modified to include a range of conditions in which successful outcomes were either very expected, expected, unpredictable, unexpected and very unexpected. In line with previous findings (e.g., Holroyd et al., 2009) and a strict interpretation of the Rescorla-Wagner learning rule (Rescorla & Wagner, 1972), one of our hypotheses was that there would be a linear relationship between the amplitude of the reward positivity and expectancy. However, our alternative hypothesis was that we would find a non-linear relationship between the amplitude of the reward positivity and expectancy – a result in congruence with biological research (e.g., a sigmoidal relationship). Furthermore, we sought to determine whether the broadened range of expectancies would cause a broadened range of changes in behaviour. Thus, in line with Holroyd and Krigolson (2007), we hypothesized that the behavioural adaptations as measured by changes in response times following positive and negative feedback would be larger following incorrect trials than correct trials and would follow the same trend as the reward positivity across expectancies.

2. Methods

2.1. Participants

Twenty undergraduate students (10 female, mean age: 22) from Dalhousie University participated in the experiment. All participants had normal or corrected-to-normal vision, no known neurological impairments, and volunteered for extra course credit in a psychology course. The data of two participants were removed from post-experiment analyses – due to an excessive number of artifacts in the EEG data of one subject and to errors in the experimental procedure for the other. All participants provided informed consent approved by the Health Sciences Research Ethics Board at Dalhousie University, and the study followed ethical standards as prescribed in the 1964 Declaration of Helsinki.

2.2. Apparatus and procedure

Participants were comfortably seated in a soundproof room in front of a computer monitor and used a standard USB gamepad to perform a modified time estimation task (Miltner, Braun, & Coles, 1997) written in MATLAB (Version 8.42, Mathworks, Natick, U.S.A.) using the Psychophysics Toolbox extension (Brainard, 1997). The time estimation task has been used previously to manipulate reward expectancy (e.g., Holroyd & Krigolson, 2007). On each trial of the task, participants were asked to estimate the duration of one second. Participants were cued to