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

begin their estimation by a 50 ms auditory tone (3000 Hz) and depressed a button on the gamepad when they believed one second had elapsed. Following the participant's response, a fixation cross was centrally presented for a brief duration (500–800 ms) after which a feedback stimulus was presented for 1000 ms. The feedback stimulus was presented in light grey on a dark grey background and consisted of either a check mark for correct trials or an 'X' for incorrect trials. A trial was considered correct when the participant's response fell within a response window centered on the target estimation time. Prior to the next trial, a blank screen was presented for a brief period of time (500–800 ms).

The response window was initially set to be 1000 ms ± 100 ms (i.e., 900 ms to 1100 ms) after the auditory cue. After each correct trial the response window decreased in size and conversely after each incorrect trial the response window increased in size, which ensures that the feedback probabilities are consistent across participants. For example, in the control condition, the degree to which the response window increased or decreased after correct and incorrect performance was equal (15 ms). Although participants estimated the second-long interval with varying degrees of precision, they each reached an equilibrium consisting of roughly half correct and half incorrect trials. Further, the degree to which the response window increased or decreased was dependent on experimental condition (*very expected*, *expected*, *control*, *unexpected*, and *very unexpected*; see Table 1), which determined the difficulty of the condition and participants' expectations of success. For example, in the very expected condition the response window decreased by a small amount on correct trials, becoming only slightly more difficult, and increased by a large amount following incorrect trials, becoming much easier, resulting in participants receiving more positive feedback than negative feedback. Theoretical and actual outcome feedback proportions are provided in Table 1. At the start of each block, the response window size was initialized with the final response window size of the previous block and within each block participants only encountered trials from one experimental condition.

The experiment began with a practice block that constituted 20 trials with a two second target time and a change of response window size as in the control condition (± 15 ms) so participants could gain familiarity with the task. Participants completed two blocks of 80 trials for each of the five experimental conditions. As such, there were a total of 800 experimental trials across all five conditions per participant. The sequencing of experimental conditions was randomly counterbalanced across participants. The task lasted on average 62 min [95% confidence intervals: 61 min, 63 min].

2.3. Data acquisition

Response time (ms) and accuracy (correct or incorrect) data were recorded by the experimental program. Electroencephalographic (EEG) data from 64 electrodes that were mounted in a fitted cap with a standard 10-10 layout (ActiCAP, Brain Products GmbH, Munich, Germany) were recorded using Brain Vision Recorder software (Version 1.10, Brain Products GmbH, Munich, Germany). All electrodes were referenced to a common ground and, during recording, electrode impedances were kept below 20 kΩ. EEG data were sampled at 500 Hz, amplified (ActiCHamp, Revision 2, Brain Products GmbH, Munich, Germany), and filtered through an antialiasing low-pass filter of 8 kHz.

2.4. Data analysis

2.4.1. Behavioural analysis

For each condition (very expected, expected, control, unexpected, very unexpected) and feedback outcome (positive, negative), we computed mean response times and mean accuracies for each participant. Furthermore, we computed the absolute difference of mean change in response times following correct and incorrect trials for each condition and feedback outcome to examine whether there were changes in behaviour related to differences in reward expectancy.

2.4.2. Electroencephalographic analysis

All EEG processing was conducted in Brain Vision Analyzer (Version 2.1.1, Brain Products GmbH, Munich, Germany). For each participant and channel the continuous EEG data were first re-referenced to an average mastoid reference and were then filtered using a dual-pass phase free Butterworth filter (pass band: 0.1 Hz to 30 Hz; notch filter: 60 Hz). After this, epochs of data were extracted from the continuous EEG from 1000 ms before to 2000 ms after every event of interest. Events of interest in the present study were experiment condition (very expected, expected, control, unexpected, very unexpected) and feedback valence (positive, negative) thus yielding 10 bins of EEG data for each participant (e.g., very expected positive, very expected negative). Long (3000 ms) epochs were extracted from the continuous EEG to facilitate independent component analysis (ICA) native to Brain Vision Analyzer to identify and remove blinks and other eye movement artifacts (Luck, 2014). A restricted fast ICA with classic PCA sphering was used in which processing continued until either a convergence bound of 1.0×10^{-7} or 150 steps had been reached. Subsequent to this, a visual examination of component head maps in conjunction with an examination of the related factor loadings was used to select components to be removed to correct ocular artifact via ICA back transformation. Following from this, the EEG data were re-segmented to a shorter 800 ms interval for each event of interest (200 ms before to 600 ms after). Data were then baseline corrected using a 200 ms window prior to feedback stimulus onset and were submitted to an artifact rejection algorithm that removed segments of data that had gradients greater than 10 μV/ms or an absolute difference of more than 150 μV (segment maxima minus segment minima) within the segment. The artifact rejection algorithm resulted in a loss of 7.5% of the total EEG data, on average, for each participant.

Event-related potential waveforms were then constructed for each participant and channel by averaging the segmented EEG for each event of interest, and grand average ERP waveforms were constructed by averaging the ERPs across participants. Next, difference waveforms were constructed for each participant and channel for each level of expectancy by subtracting negative feedback ERPs from positive feedback ERPs (Luck, 2014; Table 2).

For example, the expected condition difference waveforms were

Table 1
Experimental manipulation of task difficulty. Responses were deemed correct when they occurred within a temporal window centered around the one second mark; task difficulty was manipulated as a function of how this response window shrank (made more difficult) or grew (made easier) after correct and incorrect trials, respectively. The degree of change is reported as increment correct and incorrect for each condition. Based on these increments, predictions of success for each condition are reported as correct and incorrect probability. The actual mean percentages of success are also reported for comparison.

Condition of Difficulty	Increment Correct	Increment Incorrect	Correct Probability	Incorrect Probability	Correct Actual	Incorrect Actual
Very Expected	3 ms	30 ms	90%	10%	84%	16%
Expected	3 ms	12 ms	75%	25%	71%	29%
Control	15 ms	15 ms	50%	50%	52%	48%
Unexpected	12 ms	3 ms	25%	75%	29%	71%
Very Unexpected	30 ms	3 ms	10%	90%	15%	85%

Table 2
Conditional waveform subtractions to create reward positivity difference waveforms for all conditions of expectancy.

Condition of Expectancy	Feedback Waveform Subtractions
Very Expected	Positive in very expected − Negative in very unexpected
Expected	Positive in expected − Negative in unexpected
Control	Positive in control − Negative in control
Unexpected	Positive in unexpected − Negative in expected
Very Unexpected	Positive in very unexpected − Negative in very expected

created by subtracting the ERPs to negative feedback in the unexpected correct condition from the ERPs to positive feedback in the expected correct condition. This procedure isolated the effect of feedback valence and/or the interaction of feedback valence and probability by controlling for a main effect of event probability (see Holroyd & Krigolson, 2007; Sambrook & Goslin, 2015). Finally, averaging the corresponding individual difference waveforms across all participants created five grand average difference waveforms. To determine the scalp distribution of the reward positivity, and to assess the presence of component overlap, topographic maps were created for each condition by averaging individual participant scalp topographies at the time of their respective reward positivity peaks.

For each of the five expectancy conditions (very expected, expected, control, unexpected, very unexpected) the reward positivity amplitude was measured as the maximal deflection between 200 and 350 ms in the participant average waveforms following feedback stimulus onset at channel FCz, where the peaks were maximal and in line with previous literature (Krigolson & Holroyd, 2007; Krigolson et al., 2014).

2.4.3. Statistical procedures

Statistics were conducted on accuracy to ensure that the experimental manipulation of difficulty was successful and on reaction time to determine corresponding changes in behaviour related to neural signals. A one-way repeated measures ANOVA was conducted on accuracy rates across conditions and followed up with a trend analysis to describe the relationship. These same statistical procedures were carried out on reaction time scores. Statistics on neural data focused on the reward positivity. First, we checked for differences in the peak latency of the reward positivity across conditions with a one-way repeated measures ANOVA. Second, a one-way repeated measures ANOVA was used to determine whether the amplitude of the reward positivity changed across conditions. Repeated measures *t*-tests with a Holm correction (Holm, 1979) were then conducted to determine where the amplitude differed. Finally, a post-hoc trend analysis was conducted to determine the relationship of this change. The functions tested were sigmoidal, linear, quadratic, and cubic, and the fit was determined by variability explained (R^2).

ANOVAs, *t*-tests, and trend analyses were conducted in SPSS (Version 23, IBM Corp., Armonk, U.S.A.). The trend analysis for the reward positivity amplitude was conducted using custom code developed in MATLAB (Version 8.42, Mathworks, Natick, U.S.A.). Corrections of *t*-tests using the Holm method were performed using R Studio (Version 0.99.902, RStudio Inc., Boston, U.S.A) and R (Version 3.3.0, The R Foundation, Vienna, Austria).

3. Results

3.1. Behavioural data

A repeated measures ANOVA (condition: very expected, expected, control, unexpected, and very unexpected) with a Greenhouse-Geisser correction (assumption of sphericity was violated, $X^2(9) = 26.85$, $p = 0.002$) revealed that participants' accuracy decreased with increasing condition difficulty, $F(2.258, 38.38) = 437.60$, $p < 0.001$, $\eta_p^2 = 0.963$ (linear trend, $F(1.17) = 409.43$, $p < 0.001$, $\eta_p^2 = 0.960$; see Table 1). A subsequent repeated measures ANOVA (condition: very expected, expected, control, unexpected, and very unexpected) corrected with Greenhouse-Geisser (assumption of Sphericity was violated, $X^2(9) = 30.77$, $p < 0.001$) revealed that response times increased as the outcomes (negative feedback) became more unexpected, $F(1.84, 31.26) = 3.61$, $p = 0.042$, $\eta_p^2 = 0.175$ (quadratic trend, $F(1.17) = 5.72$, $p = 0.028$, $\eta_p^2 = 0.252$; see Fig. 1).

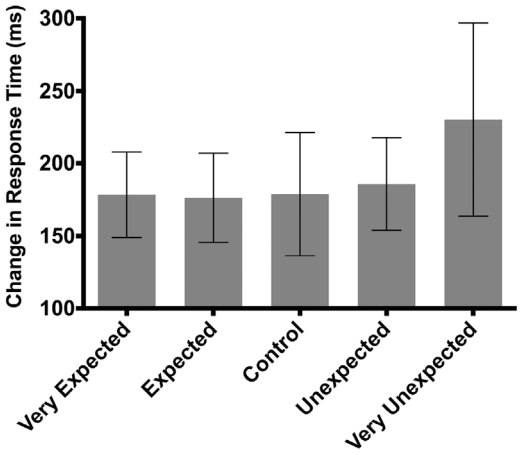

Fig. 1. Behavioural data. Change in response times following negative feedback for all levels of expectancy. Error bars indicate 95% confidence intervals.

3.2. Electroencephalographic data

The grand average difference waveforms revealed an ERP component with a timing and scalp topography consistent with the reward positivity in all of the experimental conditions (see Figs. 2 and 3). All of the difference waveforms were maximal at frontal-central areas of the scalp for all conditions. Full descriptive statistics for the reward positivity are provided in Table 3. A repeated measures ANOVA indicated that there were no significant differences between reward positivity timing across feedback expectancies, $F(4.68) = 1.65$, $p = 0.172$, $\eta_p^2 = 0.089$. The assumption of sphericity was met, $X^2(9) = 7.04$, $p = 0.636$.

A repeated measures ANOVA with a Greenhouse-Geisser correction (assumption of sphericity was violated, $X^2(9) = 22.18$, $p = 0.009$) conducted on the reward positivity amplitude revealed that the amplitude of the component was differentially modulated by experimental condition, $F(2.50, 42.47) = 14.11$, $p < 0.001$, $\eta_p^2 = 0.453$ (see Fig. 2). As predicted, the reward positivity was larger in the unexpected condition relative to the control condition, $t(17) = 4.55$, $p = 0.002$, and larger in the control condition relative to the expected condition, $t(17) = 3.20$, $p = 0.026$. Further, the size of the reward positivity for the very unexpected and the unexpected conditions did not statistically differ, $t(17) = 0.45$, $p = 0.696$, nor did the size of the reward positivity between the very expected and expected conditions, $t(17) = 0.97$, $p = 0.696$. Moreover, a post-hoc analysis revealed that a sigmoid function best fit the data ($R^2 = 0.971$) as compared to linear ($R^2 = 0.921$) and quadratic ($R^2 = 0.947$) functions. Fig. 4A presents the reward positivity amplitudes across conditions as a function of their observed accuracy for each condition. Fig. 4B presents the difference in reward positivity amplitude between each pair of conditions that are closest in probability (error bars indicate 95% confidence intervals; Cummings, 2013).

4. Discussion

Supporting previous work, the current research demonstrates that

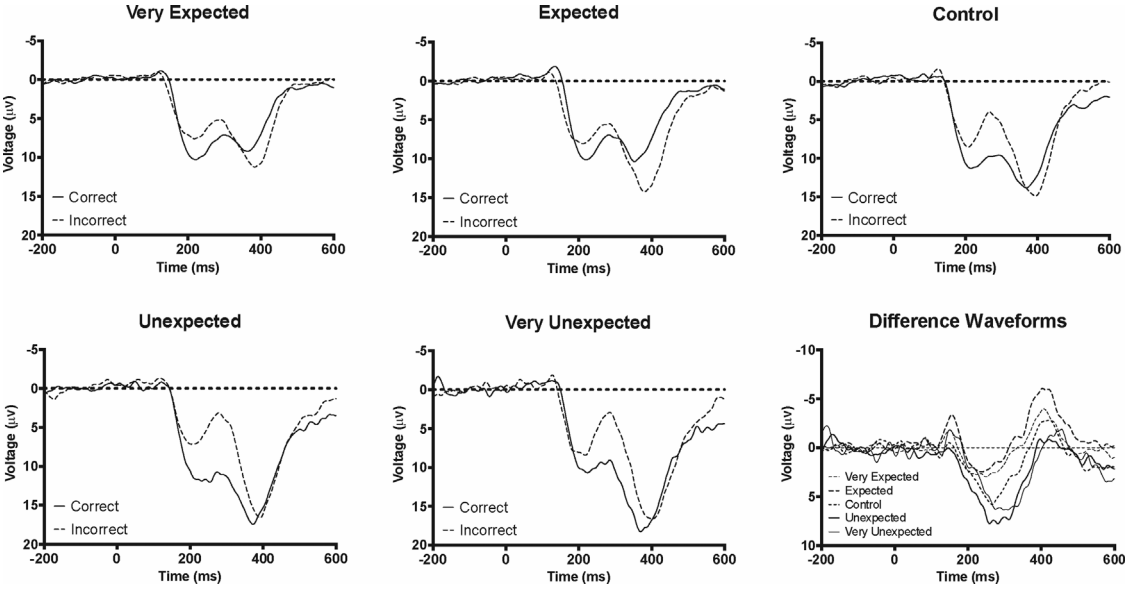

Fig. 2. Conditional waveforms and difference waveforms for all conditions of expectancy at channel FCz.

the amplitude of the reward positivity scales in magnitude with expectancy (e.g., Holroyd & Krigolson, 2007; Holroyd et al., 2009; Sambrook & Goslin, 2015). Specifically, we found that the amplitude of the reward positivity increased from the expected to the control condition, and from the control to the unexpected condition, in line with the predictions of reinforcement learning theory (i.e., Rescorla & Wagner, 1972). We also demonstrate that the amplitude of the reward positivity was not sensitive to more extreme differences in expectancy. We found that the reward positivity amplitude increased in a sigmoidal fashion as a function of unexpectedness. These findings are in contrast with theoretical accounts that state that the relationship between prediction error amplitude and expectancy is linear (e.g., Rescorla & Wagner, 1972).

Table 3
Reward positivity peak voltages and peak times for all conditions of expectancy with 95% confidence intervals at channel FCz.

Condition of Expectancy	Peak Amplitude (μV)	95% Confidence Intervals (μV)		Peak Time (ms)	95% Confidence Intervals (ms)	
Very Expected	5.3	3.75	6.82	267	243	290
Expected	4.5	2.65	6.36	263	242	283
Control	7.6	5.36	9.77	268	252	284
Unexpected	11.2	8.21	14.13	281	265	297
Very Unexpected	11.9	8.80	15.06	290	272	309

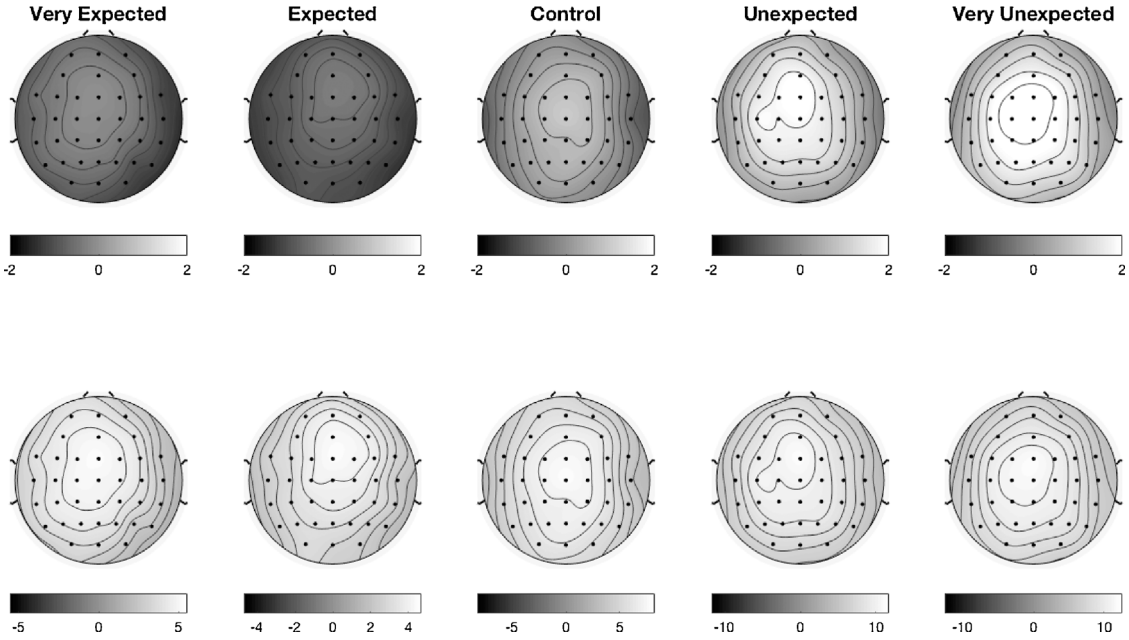

Fig. 3. Topographic maps for all conditions of expectancy. Top: Standardized topographic maps on a constant scale of voltage to demonstrate the reward positivity across conditions. Bottom: Topographic maps with different individual scales of voltage to demonstrate a scalp topography consistent with the reward positivity in each condition of expectancy. Topographic contour lines indicate a step of activity to demonstrate the spread of activity across the scalp. Each map has six contour lines equally spaced between the maximum and minimum voltage of activity on the corresponding plot. The central contour line indicates that activity is strongest at frontal-central regions of the scalp, including electrode FCz — where the reward positivity is typically analyzed.

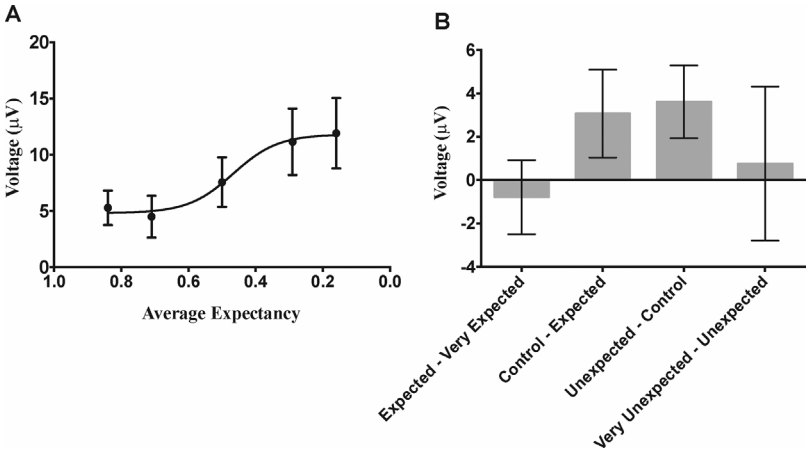

Fig. 4. Reward positivity amplitude peaks and comparisons at channel FCz. A: Peak amplitudes of the reward positivity across expectancies. Expectancies were transformed into a continuous scale by averaging the actual proportion of correct and incorrect rates for each condition of expectancy. B: Differences of reward positivity amplitudes between conditions with 95% confidence intervals.

This relationship may instead reflect the biological principles of the dopamine system (Fiorillo et al., 2003; Schultz, 2016; Schultz et al., 2015; Stauffer et al., 2016; Stauffer et al., 2014). Although it is beyond the scope of the present manuscript to provide a full background on the mechanisms underlying the reward positivity, one prominent account relates its production to phasic dopamine activity. According to the reinforcement learning (RL) theory of the reward positivity, the reward positivity reflects the impact of a dopaminergic prediction-error signal on anterior cingulate cortex (Holroyd & Coles, 2002; Holroyd & Yeung, 2012; Holroyd & McClure, 2015). Specifically, the RL theory of the reward positivity posits that anterior cingulate cortex, the basal ganglia, and the midbrain dopamine system comprise an RL system within the human midbrain and medial-frontal cortex. The basal ganglia compute a prediction error when feedback is received, which the midbrain dopamine system then conveys to anterior cingulate cortex (and other regions) to optimize behaviour. On this view, the reward positivity is the observable EEG correlate of the impact of the dopaminergic signal on anterior cingulate cortex (Holroyd & Coles, 2002; Holroyd, 2013, Holroyd & Yeung, 2012; Holroyd & McClure, 2015; Krigolson et al., 2014; Krigolson, Pierce, Tanaka, & Holroyd, 2009).

If phasic changes in dopamine influence the amplitude of the reward positivity, then factors affecting the dopamine system (such as reward magnitude and expectancy) may also affect the reward positivity. Schultz and others have provided empirical evidence that dopamine prediction-error signals change non-linearly with both reward magnitude and reward expectancy (Fiorillo et al., 2003; Schultz, 2016; Schultz et al., 2015; Stauffer et al., 2016; Stauffer et al., 2014). For example, reward expectancy has been shown to affect phasic activation of monkey dopamine neurons. Fiorillo et al. (2003) held rewards constant and observed that dopamine prediction errors scaled monotonically to reward expectedness: greater prediction errors for more unexpected rewards. Along with Stauffer et al.'s (2014) discovery of a utility function within monkey midbrain, the existence of a similar function for outcome probabilities is plausible. Such a function (relating reward expectancy to dopamine prediction errors) would provide an explanation for our reward positivity data showing no change in reward positivity at extreme levels of expectancy. Additional studies involving rewards with multiple magnitudes would determine if the reward positivity responds to reward magnitude the same way that it responds to reward expectancy here. If the relationship between reward positivity and reward magnitude resembles a utility function, it would imply that humans might also learn via a dopamine utility function (Schultz, 2016; Schultz et al., 2015; Stauffer et al., 2016; Stauffer et al., 2014).

An alternative explanation of these findings is that changes in difficulty in the extreme expectancy conditions were harder to detect than changes in the moderate expectancy conditions. Specifically, there was an average difference of 21% chance of success between the moderate conditions (e.g., between the control condition and the expected condition), yet only a difference of 13.5% chance of success between the extreme conditions (e.g., between the expected condition and very expected condition). This may have consequences on one's expectations in that the precision to which humans can distinguish between success rates may be limited. Perhaps performance differences between the extreme conditions were too small for the participants to detect and so their expectations of success did not differ. This would indicate that humans broadly generalize across expectations: if the probabilities of two events are similar enough, they are perceived as being equal. Future research could address this by collecting self-report data from participants as to their perceived likelihood of succeeding within each condition.

In sum, our findings support the claim that the expectancy of outcomes differentially modulates the reward positivity. We demonstrated that reward positivity amplitude increased between the expected, control, and unexpected outcomes. Importantly, we provide novel evidence that the neural systems that underlie human reward processing may adhere to biological principles in that there is a sigmoid relationship between the reward positivity and the unexpectedness of an event. These data suggest that while the neural computations that underlie reward processing in general follow reinforcement learning theory (e.g., Rescorla & Wagner, 1972), more accurate models of human learning should incorporate lower and upper boundaries of expectancy violations.

Funding

This work was supported by the National Science and Engineering Research Council of Canada [grant number RGPIN 2016-0943].

Acknowledgements

The neuroeconomics laboratory would like to acknowledge support from the Natural Sciences and Engineering Research Council of Canada (Olave Krigolson: Discovery: RGPIN 2016-0943).

References

Amiez, C., Joseph, J. P., & Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *The European Journal Of Neuroscience, 21*(12), 3447–3452.

Baldi, E., & Bucherelli, C. (2005). The inverted u-shaped dose-effect relationships in learning and memory: Modulation of arousal and consolidation. *Nonlinearity in Biology, Toxicology, Medicine, 3*(1) [nonlin-003].

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk (L. sommer, trans.). *Econometrica, 22*(1), 23. http://dx.doi.org/10.2307/1909829 [Original work published 1738].

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of Neurophysiology, 97*(4), 3036–3045. http://dx.doi.org/10.1152/jn.01211.2006.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307*(5712), 1118–1121.
Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage, 49*(4), 3198–3209. http://dx.doi.org/10.1016/j.neuroimage.2009.11.080.
Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *The Journal of Neuroscience, 27*(2), 371–378. http://dx.doi.org/10.1523/JNEUROSCI.4421-06.2007.
Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage, 35*(2), 968–978. http://dx.doi.org/10.1016/j.neuroimage.2006.11.056.
Cummings, L. (2013). *Pragmatics: A multidisciplinary perspective.* New York, NY: Routledge.
Eppinger, B., Kray, J., Mock, B., & Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the ERN. *Neuropsychologia, 46*(2), 521–539. http://dx.doi.org/10.1016/j.neuropsychologia.2007.09.001.
Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *The Journal of Neuroscience, 32*(35), 12087–12092. http://dx.doi.org/10.1523/JNEUROSCI.1410-12.2012.
Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science, 299*(5614), 1898–1902. http://dx.doi.org/10.1126/science.1077349.
Foster, K. A., Kreitzer, A. C., & Regehr, W. G. (2002). Interaction of postsynaptic receptor saturation with presynaptic mechanisms produces a reliable synapse. *Neuron, 36*(6), 1115–1126.
Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-Related negativity predicts reinforcement learning and conflict biases. *Neuron, 47*(4), 495–501. http://dx.doi.org/10.1016/j.neuron.2005.06.020.
Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390.
Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology, 44*(6), 905–912. http://dx.doi.org/10.1111/j.1469-8986.2007.00567.x.
Haruno, M., & Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Networks, 19*(8), 1242–1254. http://dx.doi.org/10.1016/j.neunet.2006.06.007.
Hassall, C. D., MacLean, S., & Krigolson, O. E. (2014). Hierarchical error evaluation: The role of medial-frontal cortex in postural control. *Journal of Motor Behavior, 46*(6), 381–387. http://dx.doi.org/10.1080/00222895.2014.918021.
Hewig, J., Trippe, R., Hecht, H., Coles, M. G. H., Holroyd, C. B., & Miltner, W. H. R. (2007). Decision-making in blackjack: An electrophysiological analysis. *Cerebral Cortex, 17*(4), 865–877. http://dx.doi.org/10.1093/cercor/bhk040.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics,* 65–70.
Holroyd, C. B. (2013). Theories of anterior cingulate cortex function: Opportunity cost. *Behavioral and Brain Sciences, 36*(6), 693–694.
Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679–709. http://dx.doi.org/10.1037/0033-295X.109.4.679.
Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology, 44*(6), 913–917.
Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review, 122*(1), 54.
Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences, 16*(2), 122–128.
Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport, 14*(18), 2481–2484. http://dx.doi.org/10.1097/01.wnr.0000099601.41403.a5.
Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology, 45*(5), 688–697. http://dx.doi.org/10.1111/j.1469-8986.2008.00668.x.
Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., & Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation Cognitive, Affective. *& Behavioral Neuroscience, 9*(1), 59–70. http://dx.doi.org/10.3758/CABN.9.1.59.
Holroyd, C. B., Krigolson, O. E., & Lee, S. (2011). Reward positivity elicited by predictive cues. *Neuroreport, 22*(5), 249–252. http://dx.doi.org/10.1097/WNR.0b013e328345441d.
Jessup, R. K., Busemeyer, J. R., & Brown, J. W. (2010). Error effects in anterior cingulate cortex reverse when error likelihood is high. *The Journal of Neuroscience, 30*(9), 3467–3472. http://dx.doi.org/10.1523/JNEUROSCI.4130-09.2010.
Kerr, D. S., Huggett, A. M., & Abraham, W. C. (1994). Modulation of hippocampal long-term potentiation and long-term depression by corticosteroid receptor activation. *Psychobiology, 22*(2), 123–133.
Kreussel, L., Hewig, J., Kretschmer, N., Hecht, H., Coles, M. G. H., & Miltner, W. H. R. (2012). The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity. *Psychophysiology, 49*(2), 207–219. http://dx.doi.org/10.1111/j.1469-8986.2011.01291.x.
Krigolson, O. E., & Holroyd, C. B. (2007). Predictive information and error processing: The role of medial-frontal cortex during motor control. *Psychophysiology, 44*(4), 586–595.
Krigolson, O. E., Pierce, L., Tanaka, J., & Holroyd, C. B. (2009). Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *Journal of Cognitive Neuroscience, 21*(9), 1834–1841.
Krigolson, O. E., Pierce, L. J., Holroyd, C. B., Tanaka, J. W., & Baijal, S. (2011). Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *Annals of Neurosciences, 18*(3), 113–114. http://dx.doi.org/10.5214/ans.0972.7531.1118307.
Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2014). How we learn to make decisions: Rapid propagation of reinforcement learning prediction errors in humans. *Journal of Cognitive Neuroscience, 26*(3), 635–644. http://dx.doi.org/10.1162/jocn_a_00509.
Liao, Y., Gramann, K., Feng, W., DeáK, G. O., & Li, H. (2011). This ought to be good: Brain activity accompanying positive and negative expectations and outcomes. *Psychophysiology, 48*(10), 1412–1419. http://dx.doi.org/10.1111/j.1469-8986.2011.01205.x.
Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, MA: MIT Press2014.
Martin, L. E., & Potts, G. F. (2011). Medial frontal event-related potentials and reward prediction: Do responses matter? *Brain and Cognition, 77*(1), 128–134. http://dx.doi.org/10.1016/j.bandc.2011.04.001.
Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory, Vol. 1.* New York: Oxford university press.
Matsumoto, K., Suzuki, W., & Tanaka, K. (2003). Neuronal correlates of goal-Based motor selection in the prefrontal cortex. *Science, 301*(5630), 229–232.
Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience, 10*(5), 647–656. http://dx.doi.org/10.1038/nn1890.
Mill, J. S. (1863). *Utilitarianism* (1st ed.). London: Parker, Son & Bourn, West Strand.
Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a generic neural system for error detection. *Journal of Cognitive Neuroscience, 9*(6), 788–798. http://dx.doi.org/10.1162/jocn.1997.9.6.788.
Morris, S. E., Heerey, E. A., Gold, J. M., & Holroyd, C. B. (2008). Learning-related changes in brain activity following errors and performance feedback in schizophrenia. *Schizophrenia Research, 99*(1-3), 274–285. http://dx.doi.org/10.1016/j.schres.2007.08.027.
Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G. H., Holroyd, C. B., Kok, A., & van der Molen, M. W. (2002). A computational account of altered error processing in older age: Dopamine and the error-related negativity Cognitive, Affective. *& Behavioral Neuroscience, 2*(1), 19–36. http://dx.doi.org/10.3758/CABN.2.1.19.
Nieuwenhuis, S., Heslenfeld, D. J., Alting von Geusau, N. J., Mars, R. B., Holroyd, C. B., & Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage, 25*(4), 1302–1309. http://dx.doi.org/10.1016/j.neuroimage.2004.12.043.
Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *The Journal of Neuroscience, 32*(2), 551–562. http://dx.doi.org/10.1523/JNEUROSCI.5498-10.2012.
O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*(5669), 452–454.
Ohira, H., Ichikawa, N., Nomura, M., Isowa, T., Kimura, K., Kanayama, N., et al. (2010). Brain and autonomic association accompanying stochastic decision-making. *NeuroImage, 49*(1), 1024–1037. http://dx.doi.org/10.1016/j.neuroimage.2009.07.060.
Pfabigan, D. M., Alexopoulos, J., Bauer, H., & Sailer, U. (2011). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology, 48*(5), 656–664. http://dx.doi.org/10.1111/j.1469-8986.2010.01136.x.
Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006). When things are better or worse than expected: The medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience, 18*(7), 1112–1119. http://dx.doi.org/10.1162/jocn.2006.18.7.1112.
Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology, 52*(4), 449–459. http://dx.doi.org/10.1111/psyp.12370.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. E. Prokasy (Eds.). *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience, 17*(11), 1607–1612. http://dx.doi.org/10.1038/nn.3832.
Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin, 141*(1), 213.
Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.
Schultz, W., Tremblay, L., & Hollerman, J. R. (1998). Reward prediction in primate basal ganglia and frontal cortex. *Neuropharmacology, 37*(4–5), 421–429. http://dx.doi.org/10.1016/S0028-3908(98)00071-9.
Schultz, W., Carelli, R. M., & Wightman, R. M. (2015). Phasic dopamine signals: from subjective reward value to formal economic utility. *Current Opinion in Behavioral Sciences, 5*, 147–154.
Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience, 17*, 183–195.
Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: Single neuronal signals related to degree of reward expectancy. *Science, 296*(5573), 1709–1711.
Silvetti, M., Seurinck, R., & Verguts, T. (2013). Value and prediction error estimation

account for volatility effects in ACC: A model-based fMRI study. *Cortex, 49*(6), 1627–1635. http://dx.doi.org/10.1016/j.cortex.2012.05.008.

Stauffer, W. R., Lak, A., & Schultz, W. (2014). Dopamine reward prediction error responses reflect marginal utility. *Current Biology, 24*(21), 2491–2500.

Stauffer, W. R., Lak, A., Kobayashi, S., & Schultz, W. (2016). Components and characteristics of the dopamine reward utility signal. *Journal of Comparative Neurology, 524*, 1699–1711.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel, & J. Moore (Eds.). *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497–537). MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction, Vol. 1*Cambridge: MIT press No. 1.

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., & Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience, 7*(8), 887–893. http://dx.doi.org/10.1038/nn1279.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Human neural learning depends on reward prediction errors in the blocking paradigm. *Journal of Neurophysiology, 95*(1), 301–310. http://dx.doi.org/10.1152/jn.00762.2005.

Walsh, M. M., & Anderson, J. R. (2011). Modulation of the feedback-related negativity by instruction and experience. *Proceedings of the National Academy of Sciences of the United States of America, 108*(47), 19048–19053. http://dx.doi.org/10.1073/pnas.1117189108.

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews, 36*(8), 1870–1884. http://dx.doi.org/10.1016/j.neubiorev.2012.05.008.

Wessel, J. R., Danielmeier, C., Morton, J. B., & Ullsperger, M. (2012). Surprise and error: Common neuronal architecture for the processing of errors and novelty. *The Journal of Neuroscience, 32*(22), 7528–7537. http://dx.doi.org/10.1523/JNEUROSCI.6352-11.2012.