

Feedback Related Negativity: Reward Prediction Error or Salience
Prediction Error?

by

Sepideh Heydari
BSc., from Shahid Beheshti University, 2011

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

© Sepideh Heydari, 2015
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy
or other means, without the permission of the author.

Supervisory Committee

Feedback Related Negativity: Reward Prediction Error or Salience
Prediction Error?

by

Sepideh Heydari
BSc., from Shahid Beheshti University, 2011

Supervisory Committee

Dr. Clay B. Holroyd, Department of Psychology
Supervisor

Dr. E. Paul Zehr, Neuroscience, Division of Medical Sciences
Outside Member

Dr. James Tanaka, Department of Psychology
Academic Unit Member

Abstract

Supervisory Committee

Dr. Clay B. Holroyd, Department of Psychology

Supervisor

Dr. E. Paul Zehr, Neuroscience, Division of Medical Sciences

Outside Member

Dr. James Tanaka, Department of Psychology

Academic Unit Member

The reward positivity is a component of the human event-related brain potential (ERP) elicited by feedback stimuli in trial-and-error learning and guessing tasks. A prominent theory holds that the reward positivity reflects a reward prediction error that is differentially sensitive to the valence of the outcomes, namely, larger for unexpected positive events relative to unexpected negative events (Holroyd & Coles, 2002). Although the theory has found substantial empirical support, most of these studies have utilized either monetary or performance feedback to test the hypothesis. However, in apparent contradiction to the theory, a recent study found that unexpected physical punishments (a shock to the finger) also elicit the reward positivity (Talmi, Atkinson, & El-Dereby, 2013). Accordingly, these investigators argued that this ERP component reflects a salience prediction error rather than a reward prediction error. To investigate this finding further, I adapted the task paradigm by Talmi and colleagues to a more standard guessing task often used to investigate the reward positivity. Participants navigated a virtual T-maze and received feedback on each trial under two conditions. In a reward condition the feedback indicated that they would either receive a monetary reward or not for their performance on that trial. In a punishment condition the feedback indicated that they would receive a small shock or not at the end of the trial. I found that the feedback stimuli elicited a typical reward positivity in the reward condition and an apparently delayed reward positivity in the punishment condition. Importantly, this signal was more positive to the stimuli that predicted the omission of a possible punishment relative to stimuli that predicted a forthcoming punishment, which is inconsistent with the salience hypothesis.

Table of Contents

| | |
|---|-----|
| Supervisory Committee | ii |
| Abstract | iii |
| Table of Contents | iv |
| List of Figures | v |
| Acknowledgments..... | vi |
| Dedication | vii |
| Chapter 1: Introduction and Background..... | 1 |
| <i>Participants:</i> | 6 |
| <i>Procedure</i> | 6 |
| <i>Task</i> | 7 |
| <i>Data acquisition</i> | 9 |
| <i>Data analysis</i> | 10 |
| Results..... | 11 |
| <i>Behavioral analysis</i> | 11 |
| <i>Reward Positivity</i> | 12 |
| Discussion..... | 20 |
| Footnotes..... | 25 |
| Bibliography | 26 |

List of Figures

| | |
|---|----|
| <i>Figure 1.</i> Virtual T-maze. Top: Task images presented to participants for an example trial. Bottom: associated timings..... | 9 |
| <i>Figure 2.</i> Mean reaction times..... | 15 |
| <i>Figure 3.</i> Reward positivity amplitude..... | 16 |
| <i>Figure 4.</i> Reward positivity for the (A) reward and (B) punishment conditions..... | 17 |
| <i>Figure 5.</i> Reward Positivity scalp distributions in (A) the reward condition and (B) the punishment condition..... | 18 |
| <i>Figure 6.</i> Scalp distribution of the difference wave in the punishment condition during 400-500 ms (A) and 200-300 ms (B) following feedback onset..... | 19 |

Acknowledgments

I would like to acknowledge with gratitude the support of my supervisor, Dr. Clay Holroyd, who kept me on track with words of enthusiasm and encouragement, and gave generously of his time and vast knowledge to guide me throughout this project and my MSc program in general. I am sincerely grateful to him for his understanding and patience, and for having faith in me more than I had in myself.

I am indebted to Dr. Deborah Talmi who invited me to the School of Psychological Science at the University of Manchester and kindly agreed to supervise me during the course of this experiment. She made this collaboration possible and made the whole learning experience worthwhile. I would also like to thank Dr. Wael El-Deredy who gave me generous access to the EEG lab at the University of Manchester, and made my code compatible with their hardware.

Thanks go to members of the Emotional Cognition Laboratory at the University of Manchester, especially Oana and Emily who familiarized me with the lab, and my fellow graduate students in the Learning and Cognitive Control Laboratory at the University of Victoria for maintaining a comfortable and productive working environment. I notably feel appreciative of Danesh Shahnazian who inspired me by his love of learning, his understanding of my struggles, and his scholarly advice.

Words cannot express my feeling of gratitude and pride for having the supportive and loving family that I have: my parents whose love helped me go farther than I thought I could go; my brothers, whose words of encouragement came to the rescue when I needed them the most; and my sister, to whom I have always looked up to. Finally, I am grateful to my sincere friends Sammy and Michaela for making my stay in the U.K. memorable and Reihane for patiently listening to me when I felt overwhelmed with my work.

Dedication

To my parents

and

Sam.

Chapter 1: Introduction and Background

Reinforcement learning (RL) theory describes a set of computational principles that adapt action selection and decision making processes according to received reward and penalties (Sutton & Barto, 1998). In recent years the neural mechanisms underlying these principles have been intensively studied. In particular, the midbrain dopamine system has long been associated with reinforcement learning (Wise, Spindler, de Wit, & Gerberg, 1978). Seminal work by Wolfram Schultz and colleagues indicated that dopamine neurons encode a particular type of reinforcement learning signal called a reward prediction error (RPE). On this view, brief increases in the firing rate of dopamine neurons indicate that ongoing events are “better than expected”, and brief decreases in dopamine neuron firing rate indicate that ongoing events are “worse than expected”. These RPE signals are utilized by the neural targets of the dopamine system to improve behavior according to principles of reinforcement learning (Barto, 1995; Schultz, Dayan, & Montague, 1997).

Further, converging evidence from functional magnetic resonance imaging (fMRI), electroencephalogram (EEG), and animal neurophysiology studies have suggested a role for anterior cingulate cortex (ACC) in learning from rewards and punishments (Holroyd and Yeung, 2012; Holroyd & McClure, 2015). In particular, it has been proposed that ACC utilizes phasic dopamine signals for the purpose of adaptive decision making, and that the impact of the dopamine signals on ACC activity manifests at the human scalp as a component of the event-related brain potential (ERP) called the reward positivity (Holroyd & Coles, 2002). In human trial-and-error learning or guessing tasks, feedback stimuli indicating correct, rewarding or otherwise positive outcomes elicit a relatively

positive deflection in the ERP at about 250 ms post-feedback, whereas feedback stimuli indicating error, non-rewarding or otherwise negative outcomes elicit a negative deflection in the ERP at this time (Miltner, Braun, & Coles, 1997). The reward positivity is obtained by subtracting the ERP elicited by positive feedback from the ERP elicited by negative feedback, which is distributed over frontal areas of the scalp. Although this difference was conventionally believed to be driven by an observed negativity after the presentation of error feedback, more recent studies have found that it is impacted more by a positive deflection following correct feedback; accordingly, the name has been changed from “feedback-error related negativity” to reward positivity (Holroyd, Pakzad-Vaezi, & Krigolson, 2008; see also Proudfit, 2015).

Consistent with the theory that relates the reward positivity to ACC and dopamine (Holroyd & Coles, 2002), source localization of the difference wave suggests generation in ACC (Miltner et al., 1997). Further, substantial evidence indicates that the reward positivity codes for a dopamine-like “signed” RPE signal, which manifests as a specific interaction between reward valence and reward probability: the size of the difference between unexpected negative and positive outcomes is larger than the difference between expected negative and positive outcomes (e.g., Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Holroyd & Krigolson, 2007; Holroyd, Krigolson, Baker, Lee, & Gibson, 2009). By contrast, competing accounts of this ERP component suggest that it reflects an unsigned prediction error or a “surprise” signal (Hauser, Iannaccone, Stämpfli, Drechsler, Brandeis, Walitza, et al., 2014; Oliveira, McDonald, & Goodman, 2007; Talmi, Atkinson, & El-Dereby, 2013; Talmi, Fuentemilla, Litvak, Duzel, & Dolan, 2012; Ferdinand, Mecklinger, Kray, & Gehring, 2012). On this account, the reward positivity is larger for

unexpected feedback stimuli, irrespective of the valence of the feedback. However, a recent meta-analysis has validated the proposal that this ERP component in fact reflects a (signed) RPE signal (Sambrook & Goslin, 2015).

Reward positivity amplitude is also context dependent. Previous studies have shown that reward positivity amplitude depends on the subjective value of possible outcomes rather than on their objective values. The reward positivity is produced by a system that associates reward and punishment with possible outcomes relative to other possible outcomes in a given task (Nieuwenhuis, Slagter, et al., 2005). For example, in one condition of an experiment participants either could win or not win money on each trial, and in a different condition they could either lose or not lose money on each trial (Holroyd, Larsen, & Cohen, 2004). The reward positivity was observed to be elicited by the best possible outcome in both conditions relative to the other possible outcomes (i.e., a win in the win condition and no loss in the loss condition), rather than by the objective value of the feedback (see also Nieuwenhuis, Heslenfeld et al., 2005; Holroyd et al., 2004).

Despite substantial evidence that the reward positivity reflects a RPE signal, this aspect of the theory has mainly been tested using monetary or performance feedback, so there is still some question about whether it generalizes to other types of rewards and punishments. In a pioneering study, Talmi and colleagues (2013) compared the sensitivity of the ERP to feedback stimuli indicating either monetary rewards or physical punishments. In this experiment, predictive stimuli in one condition indicated a monetary gain and in a different condition indicated a forthcoming shock to the finger. As expected, stimuli indicating a gain in the reward condition elicited a larger reward

positivity relative to stimuli indicating no-gain. However, in apparent contradiction to the reinforcement learning theory of the reward positivity, stimuli indicating a forthcoming shock in the punishment condition elicited a larger reward positivity compared to stimuli indicating shock omission. These signals shared similar spatiotemporal dynamics and morphologies, suggesting that they were the same ERP component. On the basis of these observations, the authors proposed that the reward positivity reflects a salience prediction error rather than a RPE. On this view, the reward positivity reflects an unsigned prediction error for outcomes that are particularly meaningful or relevant to the participant (such as a stimulus predicting forthcoming pain).

These results challenge a fundamental principle of the reward positivity theory, which is that the reward positivity should reflect a *signed* prediction error, being larger for unexpected positive outcomes relative to unexpected negative outcomes. However, some aspects of experimental design by Talmi and colleagues may explain this discrepancy. In particular, the task involved a passive, Pavlovian-type learning paradigm whereas most reward positivity studies involve a strong instrumental learning component. For example, effect sizes associated with the reward positivity are larger in tasks in which the feedback depends on the subjects' performance (e.g., Holroyd et al., 2009). Further, in their recent a meta-analysis, Sambrook and Goslin (2015) classified the reward positivity studies into three categories involving increasing levels of subjective control over the outcome (passive, guessing, and rule implementation), ranging from no control to complete control over the outcome. They found that reward positivity amplitude was largest when participants exerted the most control over the outcome, that is, when the feedback depended on their behavior. Lastly, Talmi and colleagues utilized several

different factors of magnitude, probability, delivery, and type of outcome in their experiment, yielding to a total of 16 ERPs across the reward and punishment conditions. Yet the reward positivity is normally obtained with a simple comparison between positive and negative feedback (e.g., Miltner, et al., 1997; Baker & Holroyd, 2009). Hence the complexity of their task may have obscured the reward positivity, as it has been observed that participants focus on the aspects of feedback that are most salient, which becomes ambiguous for feedback with multiple dimensions (Nieuwenhuis, Yeung, Holroyd, Schurger, & Cohen, 2004).

To address these issues, I adapted the experimental paradigm of Talmi and colleagues to a more standard reward positivity paradigm. In this version, on each trial participants choose to turn either left or right in a virtual T-maze to find a reward (or not) at the end of the selected alley (Baker & Holroyd, 2009). In the present experiment, participants navigated the T-maze in two conditions, the order of which were counterbalanced across subjects: in the reward condition they searched for a reward (10 pence) as in the standard version, and in a punishment condition they sought to avoid a punishment (a light shock to the finger). Note that in this context the outcomes for both conditions were framed as goals: either to achieve a reward, or to avoid a punishment. I expected to see the reward positivity elicited to feedback indicating delivery of a monetary reward in the reward condition and to feedback indicating the omission of a physical punishment in the punishment condition.

Method

Participants:

30 people (16 males) between 18 and 35 years of age participated in the study. All participants had normal or corrected-to-normal vision and none of them reported history of brain injury. Participants were recruited from the community in Manchester, U.K., via advertisement in a local newspaper (Manchester Evening News) and were compensated with 16 GBP for their time. An additional 4 GBP was paid to them at the end of the study as a bonus for the reward phase of the experiment. All participants gave written consent prior to data collection. Following task completion they also answered a questionnaire that asked about the subjective intensity of the pain stimulation and related questions. The study was approved by the Human Research Ethics board at the University of Victoria and the University of Manchester Research Ethics Committee and was conducted in accordance with the ethical standards prescribed in the 1964 Declaration of Helsinki.

Procedure

Participants were familiarized with the lab environment and equipment upon their entrance to the lab. They were then comfortably seated in front of a computer next to an in-house constructed transcutaneous electrical nerve stimulation (TENS) device. After cleaning the left index finger with alcohol and applying an electro-conductive gel, two Ag/AgCl ring electrodes were taped to the finger and connected to the TENS device. Prior to the task initiation, the strength of the painful stimulus to be delivered was determined individually for each subject as follows. Using the computer keyboard, participants were instructed that they should increase the stimulation intensity

incrementally, starting at level 0, until they felt it to be "painful but tolerable". Each time they pressed 'Enter' on the computer keyboard, the stimulation level was incremented by one unit and then delivered to the finger. In line with a previous study (Talmi et al., 2013), subjects were instructed to increase the stimulus intensity until it felt like "7 on a scale of 1 to 10". After the stimulation level was set, participants read an information sheet describing the experiment and then signed a consent form.

Participants performed the experiment in an electromagnetically shielded, quiet room with dim light. They sat 1 m away from a computer screen with the index and middle finger of their right hand comfortably positioned on left and right arrow keys, respectively. Verbal and written instructions were given to participants at the beginning of the experiment. Before beginning, they performed a few practice trials and answered some questions (e.g., "how much did you earn? Press 0 for nothing and 1 for 10 pence" in the reward condition, and "what did you experience? Press 0 for nothing and 1 for stimulation" in the punishment condition), to ensure that they understood the task requirements.

Task

The study used a standard reward positivity task to examine the effects of monetary reward vs. physical punishment on the ERP (Baker & Holroyd, 2009). Participants navigated through a "virtual T-Maze" by pressing either a left or right button to move either left or right in the maze, respectively. Images were displayed from a first-person perspective as shown in Figure 1.

As depicted in Figure 1, the trial began with an image of the stem of the T-maze (1000 ms), which was followed by an image of a double arrow that indicated to

participants that they should choose either the left or right alley by pressing a corresponding arrow key on the keyboard with the index and middle fingers of the right hand. Immediately upon responding, an image of the selected alley appeared on the screen (1000 ms) and then the feedback stimulus (either an apple, orange, pineapple, or banana; see below) was shown at central fixation overlain on the image of the alley for 500 ms. Last, a gray screen appeared and remained until the start of the following trial (1500 ms). During this period, the outcome was either delivered or not 500 ms after the onset of the gray screen, as described below.

The experiment was composed of a reward and a punishment condition, each of which were composed of two blocks of 40 trials, for a total of 160 trials per participant. Overall, 4 types of feedback were used throughout the experiment: reward, no reward, punishment, no punishment. Images of 4 types of fruit (i.e., apple, orange, banana, and pineapple) were used as feedback stimuli. At the beginning of each condition, participants were informed about the types of feedback and their corresponding fruit types. Feedback stimuli were systematically varied with outcome type (reward, no-reward, punishment, no-punishment) across participants. For example, in the reward condition a particular participant was told that at the end of each alley he or she would see an apple or an orange. For that participant, the apple indicated that the chosen alley contained 10 pence, and the orange indicated that the chosen alley contained nothing. Following the appearance of the reward feedback (in this case an apple) and 500 ms after the appearance of the gray screen, the participant heard a brief computer-generated beep; the gray screen then remained for another 1000 ms for a total of 1500 ms (Figure 1). Nothing happened during this interval following no-reward feedback. Participants were

instructed to maximize their earnings in the reward condition and were told that their accumulated earnings would be paid to them at the end of the experiment. In the punishment condition, the same particular participant was told that at the end of each alley he or she would see a banana or a pineapple. For that participant, the banana indicated that the chosen alley preceded the delivery of a shock, and the pineapple indicated that no shock was forthcoming. Following the appearance of the punishment feedback (in this case banana) and 500 ms after the gray screen appeared on the screen, the participant's left index finger was stimulated for 2 ms at the individual intensity level previously derived for that subject; the gray screen then remained for another 1000 ms for a total of 1500 ms. Nothing happened during this interval following no-punishment feedback. Participants were instructed to avoid the alley where the pain was delivered.

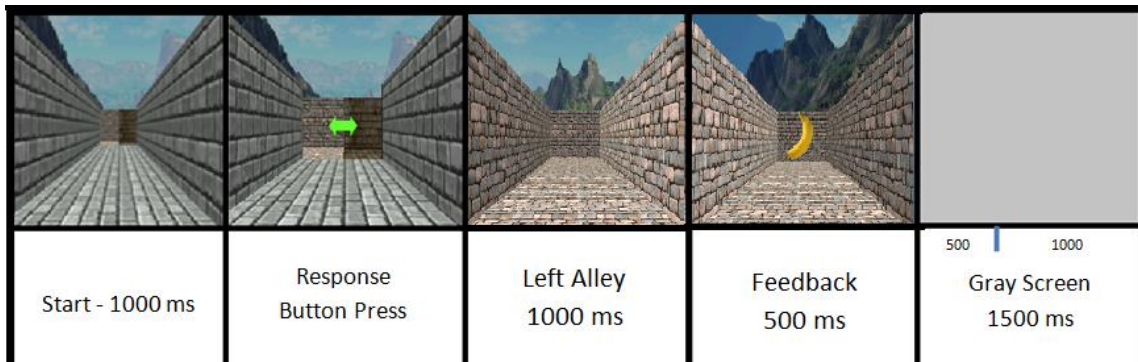


Figure 1. Virtual T-maze. Top: Task images presented to participants for an example trial. Bottom: associated timings.

Data acquisition

EEG was recorded using a Biosemi Active-Two amplifier system from 64 scalp electrodes mounted in an elastic cap (EASYCAP) according to the 10-20 system (Synamps; Neuroscan). Two additional electrodes – the common mode sense active

electrode and the driven right leg passive electrode – were used as passive and ground electrodes. Vertical (i.e. below and above the right eye) and horizontal (i.e. on temples) electrooculograms (EOG) were measured for detection of eye movement and blink artifacts. Two additional EEG electrodes were also placed over the mastoids, the data of which were later utilized for re-referencing. Impedances were maintained below 20 K Ω . EEG was recorded with bandpass filters set at 0.1-50 Hz, and a sampling rate of 512 Hz.

Data analysis

Mean reaction times were obtained for each participant and condition by averaging button press times following the onset of the double arrow on each trial. For each person and condition, trials with reaction times greater than 3 standard deviations from the mean were excluded.

EEG data were processed with Brain Vision Analyzer software (Brain Products GmbH). Raw EEG data were filtered using a Butterworth zero phase filter with a passband of 0.10--20 Hz¹. The data were segmented from 200 ms prior to and 600 ms after feedback onset. Vertical and horizontal EOG artifacts were corrected using the eye movement correction procedure algorithm developed by Gratton, Coles, & Donchin (1983). EEG data were baseline corrected for each channel and participant by subtracting the mean voltage associated with the 200 ms interval before feedback onset from the voltage at each sample during that segment. Data were then re-referenced to the average value recorded at the mastoids. ERPs were averaged separately within condition, creating for each participant and channel a reward ERP, a no-reward ERP, a punishment ERP, and a no-punishment ERP. The reward positivity was assessed with a difference wave approach by subtracting the reward ERPs from the no-reward ERPs for the reward

condition, and the no-punishment ERPs from the punishment ERPs for the punishment condition, for each participant and channel (Miltner et al., 1997; Holroyd & Krigolson, 2007). Reward positivity amplitude was determined as the average of the area under the curve for the difference waves from 240-340 ms following feedback onset at channel FCz, as prescribed in a recent meta-analysis of reward positivity studies (Sambrook and Goslin, 2015). To verify the frontal central distribution of the ERP component, scalp distributions of the average reward positivity amplitude from 240-340 ms following feedback onset were inspected for both the reward and punishment conditions.

Results

Behavioral analysis

Average stimulus intensity was 32.6 ± 23.8 (range from 3 to 230; numbers exceeding one standard deviation from the mean were excluded as outliers) and the average of the perceived intensity reported by participants was 6.5 (on a scale of 1 to 10). This level of pain is consistent with that reported in Talmi and colleagues (2013) and indicates that the stimulus was, in fact, painful.

For the reward condition, about 1% of trials were excluded as outliers for the reaction time analysis (1.7 ± 1.2 trials across subjects, range 1-7 trials; see methods). For the punishment condition, about 0.9% of trials were excluded as outliers for the reaction time analysis (1.5 ± 0.7 trials, range 1-3). A mixed design ANOVA on reaction time with order (reward first, punishment first) as a between-subject factor and condition (reward, punishment) as a within-subject factor revealed a main effect of condition ($F=7.4$, $p = 0.01$) and a significant interaction between condition and order ($F=10.7$, $p < 0.005$) (Figure 2). The main effect of order was not significant ($F=0.6$, $p > 0.05$). Post-hoc

analysis revealed that for the participants who completed the reward condition first, reaction times in the reward (0.91 ± 0.50 s) and punishment (0.83 ± 0.81 s) conditions did not significantly differ ($t = -0.3$, $p > 0.05$) (Figure 2, left). By contrast, for the participants who completed the punishment condition first, the reaction time in the punishment condition (1.31 ± 0.76 s) was significantly slower than the reaction time in the reward condition (0.83 ± 0.43 s) ($t = 4.3$, $p < 0.001$) (Figure 2, right).

Reward Positivity

On average $6.6 (\pm 4.9)$ trials (3.9 %) were excluded due to EEG artifact (range from 0 to 81 per participant; numbers exceeding two standard deviations from the mean were excluded as outliers).

A mixed design ANOVA on reward positivity amplitude with order (reward first, punishment first) and condition (reward, punishment) as factors revealed a main effect of condition ($F=43.8$, $P<0.001$), no main effect of order ($F=0.001$, $p = 0.97$), and no interaction between order and condition ($F=0.8$, $P = 0.36$) (Figure 3). Post-hoc analysis indicated that the amplitude of reward positivity in the reward condition when performed first (-5.8 ± 1.01 μV) was similar to the amplitude of the reward positivity in the reward condition when performed second (-6.7 ± 1.4 μV ; $t=-0.5$, $p>0.05$), and was higher than reward positivity amplitude in the punishment condition, both when the punishment condition was carried out first (0.7 ± 0.5 μV ; $t=3.6$, $P<0.005$) and second (-0.2 ± 1.1 μV ; $t=6.2$, $p<0.001$) (Figure 3). Finally, the amplitude of reward positivity in the punishment condition was similar irrespective of condition order ($t=0.4$, $p = 0.05$). Therefore, for subsequent analyses the data were collapsed across group (i.e., condition order). Figure 4 illustrates the feedback ERPs recorded at channel FCz and their associated difference

waves in the reward and punishment conditions. Averaged across order, the mean reward positivity amplitude of in the reward condition ($-6.3 \mu\text{V} \pm 0.8 \mu\text{V}$) was significantly different from zero ($t = 7.3, p < 0.001$). By contrast, mean reward positivity amplitude in the punishment condition ($0.2 \pm 0.6 \mu\text{V}$) was not statistically different from zero ($t = -0.3, p > 0.05$).

Figure 5 illustrates the scalp distributions of the reward positivity for both reward and punishment conditions. The reward positivity in the reward condition was maximal at channel CPz ($-6.38 \pm 0.8 \mu\text{V}$), but this value was not statistically different from reward positivity amplitude at channel FCz ($-6.30 \pm 0.8 \mu\text{V}; t=0.7, p>0.05$), consistent with a typical reward positivity (Figure 5, left panel). The reward positivity in the punishment condition was maximal at channel P10 (Figure 5, right panel). The mean reward positivity amplitude at channel P10 ($-0.4 \pm 0.7 \mu\text{V}$) was larger than the amplitude at channel FCz ($0.2 \pm 0.6 \mu\text{V}$) but not significantly different from it ($t = 0.7, p > 0.05$). The most positive area of the difference wave was located at channel FPz ($3.95 \pm 1.8 \mu\text{V}$) but was not significantly different from the amplitude at channel FCz ($t=2.03, p>0.05$). The distribution of the difference wave in the punishment condition is therefore inconsistent with its identification with the reward positivity, irrespective of the sign of the difference.

Figure 4 (right panel) also suggests differences in the ERP to positive and negative feedback in the punishment condition that occur shortly before and after the period of the reward positivity. To explore whether these differences were consistent with the reward positivity, I conducted an exploratory analysis on their scalp distributions. First, I examined the difference wave for the punishment condition from 400-500 ms post-feedback, when the difference wave reached maximum amplitude. The

difference wave at channel FCz was more negative during this period (indicating that the ERP to the no-punishment feedback was more positive than the ERP to the punishment feedback). The distribution of the difference wave was fronto-central and maximal at channel Cz ($-3.9 \pm 7.2 \mu\text{V}$), which was not statistically different from the difference wave amplitude at channel FCz ($-3.6 \pm 6.4 \mu\text{V}$; $t=-0.7$, $p > 0.05$; figure 6A). These results are consistent with the possibility that the feedback stimuli in the punishment condition elicited a delayed reward positivity.

Second, I examined the difference wave for the punishment condition during the period immediately prior to the reward positivity window, from 200-300 ms. The difference wave at channel FCz was positive-going (indicating that the ERP to no-punishment feedback was more negative than the ERP to punishment feedback), but was not statistically different from zero ($t = -1.3$, $p > 0.05$). The distribution of the difference was positively maximal at channel CP5 ($4.1 \pm 4.6 \mu\text{V}$) and negatively maximal at channel Pz ($-0.5 \pm 3.3 \mu\text{V}$; figure 6B). These scalp distributions are inconsistent with the reward positivity.

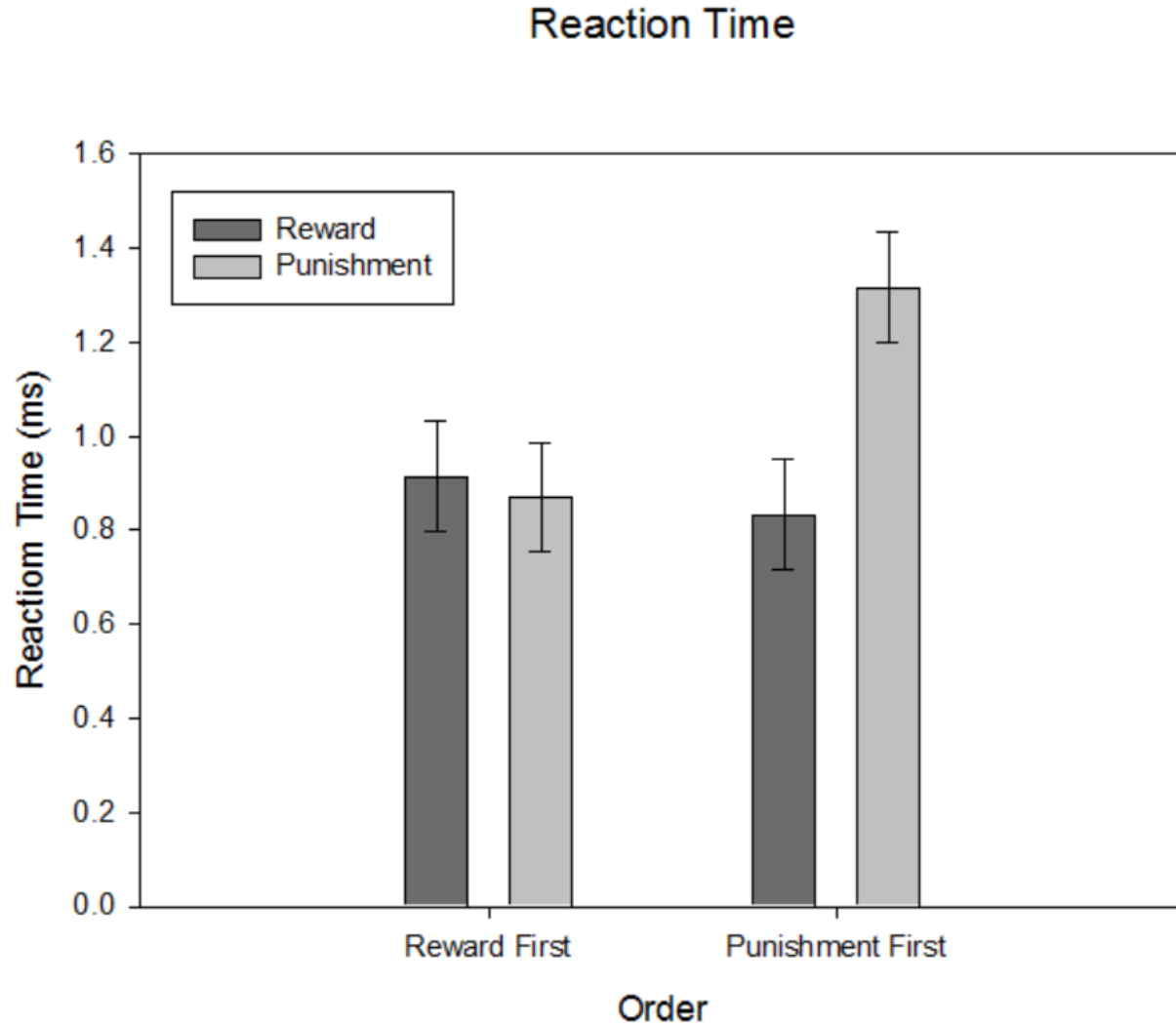


Figure 2. Mean reaction times.

Mean reaction times for the reward (dark grey) and punishment (light grey) conditions for participants who completed the reward condition first (left) and participants who completed the punishment condition first (right). Error bars are 95% within-subjects confidence intervals (Loftus & Masson, 1994).

Reward Positivity Amplitude

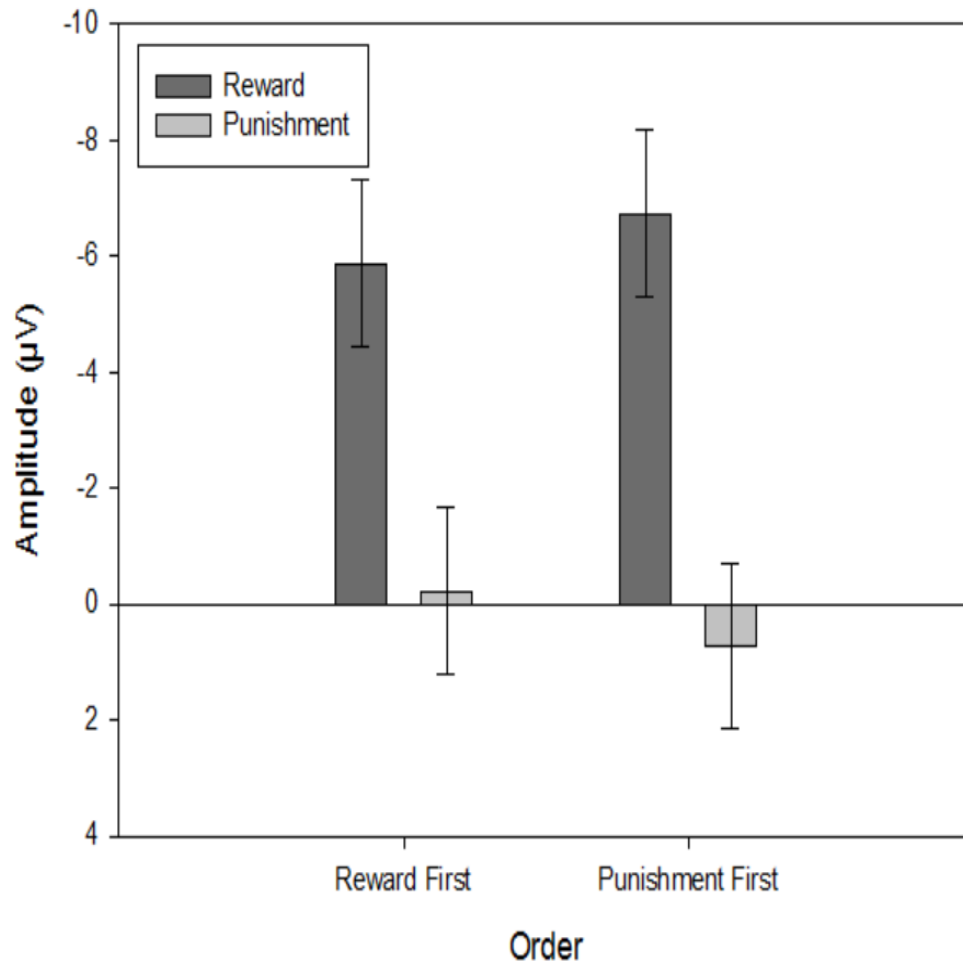


Figure 3. Reward positivity amplitude.

Reward positivity amplitude in the reward (dark grey) and punishment (light grey) conditions for participants who completed the reward condition first (left) and participants who completed the punishment condition first (right). Error bars indicate 95% within-subjects confidence intervals (Loftus & Masson, 1994).

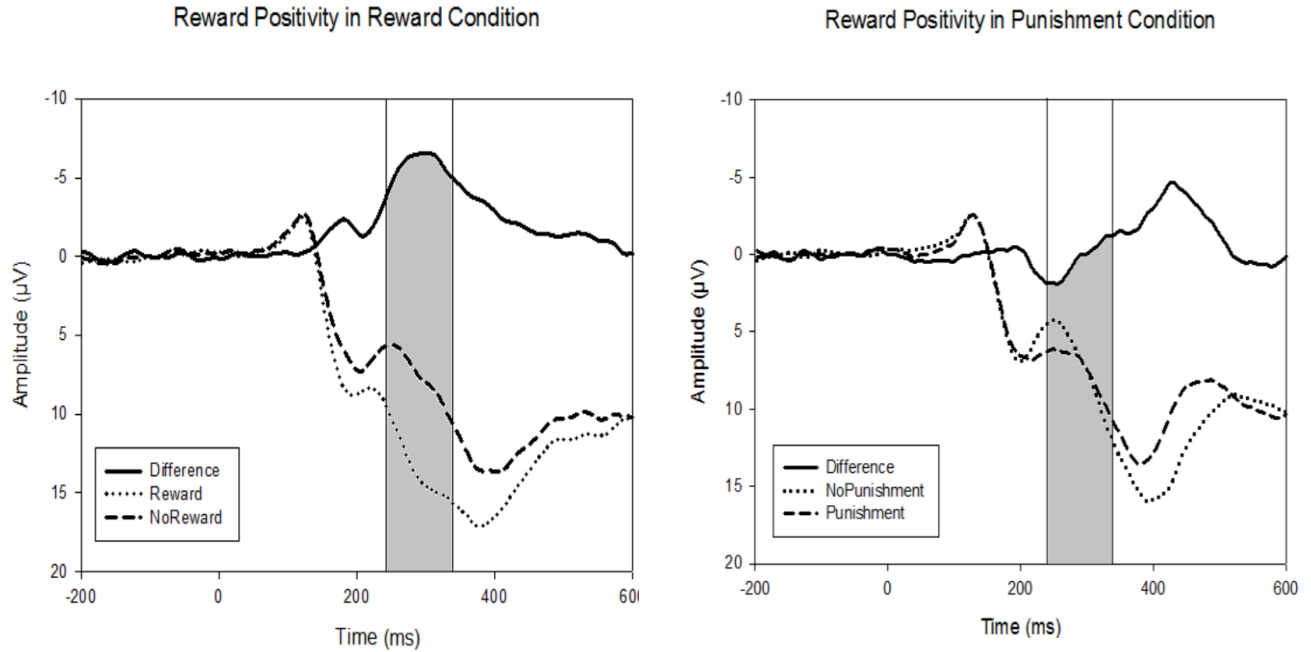


Figure 4. Reward positivity for the (A) reward and (B) punishment conditions.

Dotted lines: event-related brain potentials (ERPs) to positive feedback stimuli (i.e., reward feedback in the reward condition and no-punishment feedback in the punishment condition). Dashed lines: ERPs to negative feedback stimuli (i.e., no-reward feedback in the reward condition and punishment feedback in the punishment condition). Solid lines: difference waves associated with corresponding ERPs. Zeros on abscissae indicate time of feedback stimulus onset. Negative is plotted up by convention. Shaded area indicates the time period when the reward positivity is evaluated (Sambrook and Goslin, 2015). Data recorded at channel FCz. Note that the reward positivity is evident in the reward condition (A) but not in the punishment condition (B).

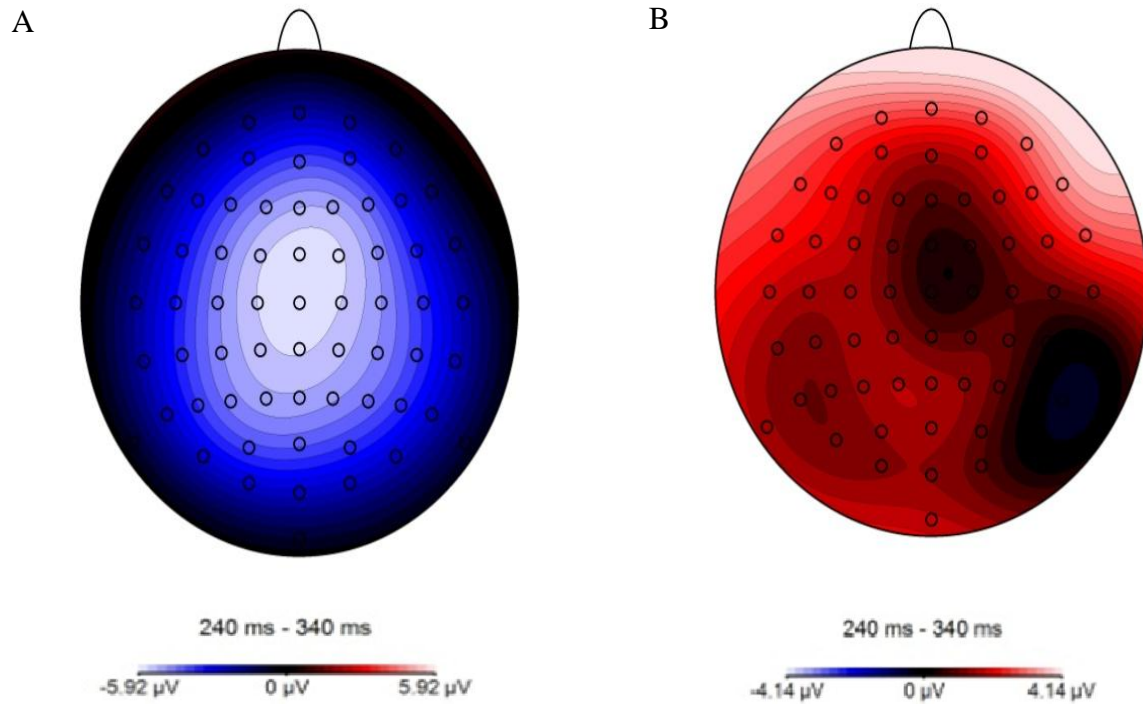


Figure 5. Reward Positivity scalp distributions in (A) the reward condition and (B) the punishment condition.

The difference wave from 240 – 340 ms post-feedback in the reward condition (left) reached maximum amplitude at frontal-central scalp locations, consistent with the reward positivity. The difference wave from 240- 340 ms post-feedback in the punishment condition (right) reached maximum amplitude at right posterior locations.

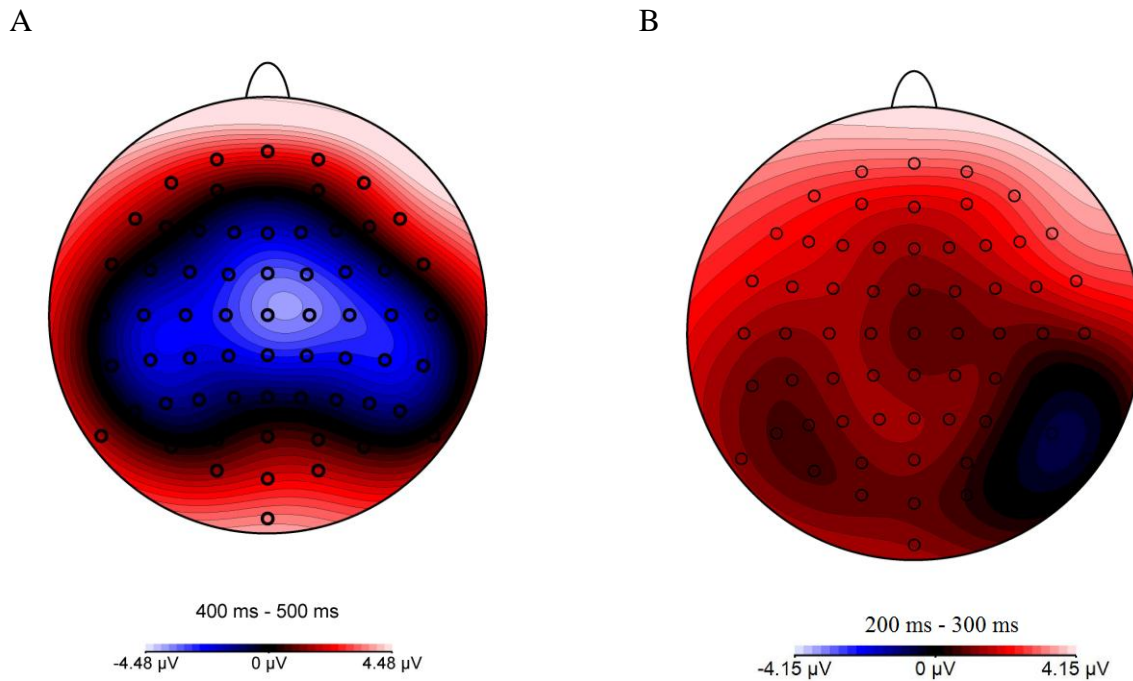


Figure 6. Scalp distribution of the difference wave in the punishment condition during 400-500 ms (A) and 200-300 ms (B) following feedback onset.

The difference wave from 400 – 500 ms post-feedback in the punishment condition (left) reached maximum amplitude at frontal-central scalp locations, consistent with the reward positivity. The difference wave from 200- 300 ms post-feedback in the punishment condition (right) reached maximum amplitude at right posterior locations.

Discussion

Substantial evidence has confirmed the proposal that reward positivity amplitude reflects a reward prediction error signal when evaluated with respect to monetary or performance feedback (Sambrook and Goslin, 2015; see also Proudfit, 2015; Walsh & Anderson, 2012). By contrast, in a previous study Talmi and colleagues (2013) found that stimuli that predicted a forthcoming punishment also elicited the reward positivity (with respect to stimuli that predicted the omission of the punishment), in an apparent violation of the reward positivity theory. In particular, given that the theory proposes that the reward positivity reflects a (signed) RPE (Holroyd & Coles, 2002), and evidence that reward positivity amplitude is context dependent (being largest for the best outcome in a range of possible outcomes; Holroyd et al., 2004), the observation of a larger reward positivity to this relatively negative outcome clashes with this theoretical position. On this basis, Talmi and colleagues proposed that this ERP component reflects a *saliency* (unsigned) prediction error as opposed to a RPE (see also Garofalo, Maier, & di Pellegrino, 2014).

Here I investigated this issue by attempting to replicate the result using a task that more closely aligns to the previous reward positivity literature. My task departed from that used by Talmi and colleagues in two important ways: task simplicity and control over outcome. First, by simplicity I mean that the task paradigm was straightforward both for the participant to follow and for data interpretation. The outcome varied according to only two factors, each with two levels – delivery (delivered/omitted), and type (reward/punishment) – yielding 4 ERPs, the minimum necessary to test the prediction of interest (“Does punishment generate a reward positivity?”). By contrast, in the study by

Talmi and colleagues the feedback varied by 4 factors – expectancy (high/low), delivery (delivered/omitted), magnitude (high/low), and type (punishment/reward) – yielding 2^4 or 16 possible outcomes. Given that participants are sensitive to the dimension of feedback that is most salient (Nieuwenhuis et al., 2004), the complexity of the task and feedback may have obscured or disrupted the standard reward positivity. Second, in the experiment by Talmi and colleagues the subjects participated passively with no requirement to actively respond to the stimuli to achieve rewards or avoid punishments. The reward positivity is observed to be reduced in task paradigms that minimize control over the outcome (Holroyd et al., 2009; Sambrook & Goslin, 2015). By contrast, in the virtual T-maze task used here the subjects were instructed to select on each trial the alley that they believed contained money in the reward condition, and to avoid a shock in the punishment condition. Although feedback in the task is delivered with 50/50 probability of reward (unbeknownst to the subjects), such that the task is associated with an intermediate level of control over the outcome (Sambrook & Goslin, 2015), the task nevertheless generates a robust reward positivity (Baker & Holroyd, 2009).

Consistent with previous findings, I found that the reward condition in the T-maze task elicited the reward positivity, which exhibited the standard spatial-temporal pattern at about 240- 340 ms observed over frontal central areas of the scalp. However, in contrast to the findings of Talmi and colleagues, the punishment condition failed to reveal a reward positivity to punishment. In fact, reward positivity amplitude in the punishment condition was not statistically different from zero μV at channel FCz, but rather exhibited a far-frontal scalp distribution that was most positive at channel FPz. These results indicate that the punishment feedback in this task did not elicit a reward positivity.

Importantly, participants rated the pain stimulus to be as subjectively painful as in Talmi and colleagues' original study, so there is no reason to believe that the stimulation in this study was any less salient than before. On this basis, I conclude that the reward positivity does not reflect a salience prediction error.

Nevertheless, the results also failed to confirm our prediction that the omission of punishment would elicit a reward positivity. Multiple considerations bear on this issue. First, there seems to be a delayed reward positivity in the punishment condition occurring between 400-500 ms following feedback onset. This time period is typically associated with the P300, a large, posteriorly distributed ERP deflection associated with other aspects of task processing (Donchin & Coles, 1988). As illustrated in figure 6, the grand average ERP in the punishment condition is more negative than the grand average ERP in the no punishment condition over central areas of the scalp, suggesting that this difference may reflect a reward prediction error signal. Second, a difference between the average ERPs in the punishment and no punishment conditions from 200-300 ms post-feedback is inconsistent with its identification as the reward positivity in terms of timing, polarity, and scalp distribution. Rather, the difference may reflect an early saliency effect reported in the meta-analysis by Sambrook and Goslin (2015), as well as the saliency effect to punishment observed by Talmi and colleagues (2013), which also occurred relatively early (about 230-270 ms after feedback onset) and was more frontally distributed (maximal at channel Fz). Taken together, these results indicate that feedback stimuli that predict a forthcoming physical punishment do not elicit the reward positivity, indicating that this ERP component does not code for stimulus salience. Rather, the absence of a potential and imminent punishment appears to elicit a reward positivity that

is delayed in time by about 150 ms, a possibility that has been observed elsewhere (e.g., Baker & Holroyd, 2011).

Why was the reward positivity delayed following feedback stimuli that indicated that a potential forthcoming punishment would not be delivered? One possibility is that participants adopted a different task strategy in the punishment condition that delayed the reward positivity, as suggested by the observed increase in reaction time for the participants who performed the punishment condition first (Figure 2). However, reward positivity amplitude did not reveal a comparable interaction of condition and valence (Figure 3), suggesting this factor does not underlie the dissociation. It may be that the mesocortical dopamine system processes the pursuit of a positive goal fundamentally differently from the avoidance of a negative outcome. Further investigations of this proposal remain for future studies.

Finally, in both the study by Talmi and colleagues and the present experiment, the outcomes differed in terms of their timing: while the punishments were immediate on each trial, the rewards were delayed until the end of the experiment when the winnings were paid out in a lump sum. Note that this difference between the conditions does not affect the conclusions of either experiment: the RPE theory of the reward positivity predicts that the predictive stimuli should elicit a reward positivity irrespective of whether the outcomes are immediate or delayed, for both punishment and reward; whereas the results of Talmi et al. appear contrary to this prediction, our present results support it.

I propose an experiment that would examine the time course of reward positivity amplitude across early and late trials in a block. This is particularly interesting because it

would allow for investigating whether each punishment indicates progress through the task. In particular, it is possible that some participants saw each punishment as indicating progress through an onerous task and thus as rewarding. This analysis was not possible in the current study due to an insufficient signal-to-noise ratio to produce a clear reward positivity for subsets of trials. Alternatively, in a different experiment participants could be instructed to collect a fixed number of punishments, or a fixed number of punishment omissions (say, 50), in order to complete the task. I predict that feedback indicating punishment will elicit the reward positivity when subjects seek the punishments in order to finish the task faster, and will be elicited by feedback indicating punishment omissions when they do the same for these feedback stimuli.

Footnotes

¹For 8 participants the EEG data at one channel (P2) was interpolated using the Hjorth Nearest Neighbors algorithm; for two other participants, the data associated with 3 and 8 channels over posterior and occipital areas) away from the region of interest over frontal cortex) were also interpolated.

Bibliography

- Baker, T. E., & Holroyd, C. B. (2009). Which way do I go? Neural activation in response to feedback and spatial processing in a Virtual T-Maze. *Cerebral Cortex*, *19*, 1708-1722.
- Baker, T. E., & Holroyd, C. B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biological psychology*, *87*(1), 25-34.
- Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215-232). Cambridge, MA: MIT Press.
- Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage*, *35*(2), 968-978.
- Donchin E., & Coles M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, pp 357-374
doi:10.1017/S0140525X00058027.
- 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.
- Garofalo, S., Maier, M. E., & di Pellegrino, G. (2014). Mediofrontal negativity signals unexpected omission of aversive events. *Scientific Reports*, *4*, 4816.
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, *55*(4), 468-484.
- Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage*, *84*, 159-168.
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *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., 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.
- Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology*, 41(2), 245-253.
- Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, 122, 54-83.
- 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.
- 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.
- Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in cognitive sciences*, 16(2), 122-128.
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic bulletin & review*, 1(4), 476-490.
- Miltner, W. H. R., Braun, C. H., & Coles, M. G. (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.
- Miltner, W. H. R., Lemke, U., Weiss, T., Holroyd, C. B., Scheffers, M. K., & Coles, M. G. H. (2003). Implementation of error-processing in the human anterior cingulate cortex: A source analysis of the magnetic equivalent of the error-related negativity. *Biological Psychology*, 64, 157-166.
- 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.
- Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J. A., Heslenfeld, D., & Holroyd, C. B. (2005). Knowing good from bad: Differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience*, 21, 3161-3168.

- Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A., & Cohen, J. D. (2004). Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex*, *14*(7), 741-747.
- Oliveira, F. T., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. *Journal of cognitive neuroscience*, *19*(12), 1994-2004.
- Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, *52*: 449–459.
doi: 10.1111/psyp.12370.
- Sambrook, T. D.; Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, Vol 141(1), 213-235. <http://dx.doi.org/10.1037/bul0000006>.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.
- Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT Press.
- Talmi, D., Atkinson, R., & El-Dereby, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *The Journal of Neuroscience*, *33*(19), 8264-8269.
- Talmi, D., Fuentemilla, L., Litvak, V., Duzel, E., & Dolan, R. J. (2012). An MEG signature corresponding to an axiomatic model of reward prediction error. *Neuroimage*, *59*(1), 635-645.
- 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.
- Wise, R. A., Spindler, J., & Gerberg, G. J. (1978). Neuroleptic-induced "anhedonia" in rats: pimozide blocks reward quality of food. *Science*, *201*(4352), 262-264.