Published in final edited form as:

Physiol Behav. 2016 July 01; 161: 24–32. doi:10.1016/j.physbeh.2016.03.034.

The temporal dynamics of reversal learning: P3 amplitude predicts valence-specific behavioral adjustment

Kayla R. Donaldson, Belel Ait, Oumeziane Sebastien, and Hlie Dan Foti

Department of Psychological Sciences, Purdue University

Abstract

Adapting behavior to dynamic stimulus-reward contingences is a core feature of reversal learning and a capacity thought to be critical to socio-emotional behavior. Impairment in reversal learning has been linked to multiple psychiatric outcomes, including depression, Parkinson's disorder, and substance abuse. A recent influential study introduced an innovative laboratory reversal-learning paradigm capable of disentangling the roles of feedback valence and expectancy. Here, we sought to use this paradigm in order to examine the time-course of reward and punishment learning using event-related potentials among a large, representative sample ($N = 101$). Three distinct phases of processing were examined: initial feedback evaluation (reward positivity, or RewP), allocation of attention (P3), and sustained processing (late positive potential, or LPP). Results indicate a differential pattern of valence and expectancy across these processing stages: the RewP was uniquely related to valence (i.e., positive vs. negative feedback), the P3 was uniquely associated with expectancy (i.e., unexpected vs. expected feedback), and the LPP was sensitive to both valence and expectancy (i.e., main effects of each, but no interaction). The link between ERP amplitudes and behavioral performance was strongest for the P3, and this association was valence-specific. Overall, these findings highlight the potential utility of the P3 as a neural marker for feedback processing in reversal-based learning and establish a foundation for future research in clinical populations.

Keywords

Reversal-learning; Flexibility; ERP; Reward; Punishment

1. Introduction

Environmental contingencies are typically not static but instead change over time, requiring the ongoing evaluation of feedback and adjustment of one's behavior in order to attain desired outcomes. Experimental studies commonly model this process using reversal learning tasks, in which stimulus-outcome contingencies are first learned and then reversed, requiring participants to update their learned associations and adapt their behavioral responses (Chase, Swainson, Durham, Benham, & Cools, 2011; Cools, Clark, Owen, & Robbins, 2002). Functional magnetic resonance imaging (fMRI) studies have shown that

reversal learning acts on the brain's dopaminergic reward pathway, including parts of the prefrontal cortex and the striatum (Cools et al., 2002; Robinson, Frank, Sahakian, & Cools, 2010). Impairment in reversal learning, meanwhile, has been implicated in multiple psychiatric and neurological disorders characterized by dopaminergic dysfunction, including substance use disorders (Patzelt, Kurth-Nelson, Lim, & MacDonald, 2014; Torres et al., 2013), major depression (Remijnse et al., 2009; Robinson, Cools, Carlisi, Sahakian, & Drevets, 2012), and Parkinson's disorder (Cools, Altamirano, & D'Esposito, 2006; Cools, Lewis, Clark, Barker, & Robbins, 2006).

Recent studies using event-related potentials (ERPs) have shed light on the time course of feedback processing in various contexts, including reversal learning. In particular, two ERP components have garnered the most attention in relation to feedback processing: the reward positivity (RewP) and P3. The RewP (known previously as the feedback negativity [FN], or feedback-related negativity [FRN]; Proudfit, 2015), is maximal at the frontocentral electrodes, peaks approximately 250–300ms following feedback, and reflects the initial, binary evaluation of outcomes as either better or worse than expected (Foti, Weinberg, Dien, & Hajcak, 2011; Holroyd, Pakzad-Vaezi, & Krigolson, 2008). Traditionally thought to represent a negative-going ERP response to unfavorable feedback (e.g., errors, monetary loss; Gehring & Willoughby, 2002; Miltner, Braun, & Coles, 1997) recent evidence suggests that RewP amplitude is likely modulated by positive feedback (e.g., correct outcomes, monetary gain; Foti et al., 2011; Holroyd et al., 2008; Proudfit, 2015). Immediately following the RewP, the P3 is maximal at parietal sites approximately 300–500 ms following stimulus onset and is a positive-going ERP component. It is thought to reflect a reallocation of attention to salient stimuli, particularly those that are task-relevant, infrequent, or unexpected (Courchesne, Hillyard, & Galambos, 1975; Donchin & Coles, 1988).

In one influential reversal learning study, a novel laboratory paradigm was developed to disentangle the role of feedback valence by fully crossing outcome valence (good vs. bad) and expectancy (expected vs. unexpected) across blocks (Cools, Altamirano, et al., 2006). Unlike most reversal learning tasks, here the participant plays the role of an observer and is required to predict the feedback resulting from the decisions of a fictional player. A series of paired stimuli are presented simultaneously, one of which is highlighted on each trial to indicate the stimulus chosen by the player. The objective for participants is to predict whether highlighted stimulus would result in a reward or punishment for the player, based on the results of previous trials. The outcome, therefore, is not contingent upon participants' responses per se, but rather on the highlighted stimulus and the current rule. Stimulus-outcome contingencies change periodically throughout the task and are indicated to participants by either unexpected reward or unexpected punishment feedback to the player (i.e., reversal trials).

In one initial study, this task effectively isolated valence-specific impairment in reversal learning in relation to Parkinson's disease: dopaminergic medication selectively impaired reversal learning from unexpected punishment among patients, whereas learning from unexpected reward was unaffected (Cools et al., 2006). A subsequent study within a healthy sample used this same task to examine ERPs involved in reversal learning. Critically, a double dissociation was observed for outcome valence and expectancy across the RewP and

P3, respectively (von Borries, Verkes, Bulten, Cools, & de Bruijn, 2013). The P3 was uniquely affected by expectancy, such that amplitudes were larger for unexpected-outcome trials. Conversely, RewP amplitude was more negative for unfavorable feedback (i.e., punishment) and more positive for favorable feedback (i.e., reward), irrespective of outcome expectancy. These studies highlight that reward and punishment learning are dissociable processes within the broader context of reversal learning, and that the RewP and P3 can be used to effectively characterize each.

An outstanding question in the literature is how neural measures of feedback processing translate to behavioral adjustment. In the aforementioned reversal learning study, associations between RewP/P3 amplitudes and performance within the reward and punishment conditions were not considered (von Borries et al., 2013). Meanwhile, in two other studies using probabilistic learning tasks, behavioral adjustment was linked specifically with the P3 and not the RewP (Chase et al., 2011; San Martín, Appelbaum, Pearson, Huettel, & Woldorff, 2013). A caveat to some of these later findings is that the ERP response was scored in a relatively broad time window (e.g., 400–800 ms; San Martín et al., 2013), thereby encompassing both the P3 and the late positive potential (LPP). Like the P3, the LPP is a positive-going ERP at centroparietal electrodes. Unlike the P3, however, the LPP is a sustained neural response, lasting throughout stimulus presentation and even up to 1,000 ms after stimulus offset (Hajcak, Dunning, & Foti, 2009; Hajcak & Olvet, 2008). Traditionally studied in emotional viewing paradigms (Codispoti, Ferrari, & Bradley, 2007; Foti, Hajcak, & Dien, 2009; Weinberg & Hajcak, 2010), the LPP is thought to capture the sustained allocation of attention to motivationally salient stimuli (Hajcak et al., 2009). In the context of reversal learning, therefore, the P3 likely reflects the initial orientation toward unexpected outcomes, and the LPP may reflect the sustained processing of motivationally salient outcomes. It is unclear, however, the extent to which previous learning studies captured modulation of the P3 or LPP per se. It is of interest whether the P3, LPP, or a combination thereof predicts effective behavioral adjustment following contingency reversals.

The present study seeks to extend these previous findings by examining ERPs during reversal learning within a large, representative sample. We use here the paradigm developed by Cools and colleagues (Chase et al., 2011; Cools, Altamirano, et al., 2006; San Martín et al., 2013) in order to examine reward and punishment learning separately, examining two key questions: First, we examined the effects of outcome valence and expectancy across three stages of feedback processing: initial evaluation (RewP), allocation of attention (P3), and sustained processing (LPP). No reversal learning study to date has quantified the P3 and LPP separately, leaving it unclear how feedback processing unfolds across these three stages and whether there is utility in considering these late parietal ERPs separately. In a previous study, outcome valence and expectancy separately modulated the RewP and P3, respectively (von Borries et al., 2013), and we expected to see the same pattern here; we also explored how each factor affected LPP amplitude. Second, we examined the links between each of these three stages and behavioral performance. In two previous studies using probabilistic learning tasks, ERP activity in the P3/LPP time window (but not the RewP) was associated with behavioral performance (Chase et al., 2011; San Martín et al., 2013). We expected that the P3/LPP in the current task would be associated with effective learning, as indicated by

lower error rate following contingency reversals. We explored whether this association would be specific to either the P3 or LPP time windows, and whether the observed brain-behavior relationships would be further specific to the reward and punishment learning conditions.

2. Methods

2.1. Participants

Data was collected from 122 participants. Twenty-one (17.2%) were excluded from analysis (3 for equipment failure, 2 for not understanding task instructions, 10 for poor behavioral performance, 6 for poor-quality ERP data), leaving 101 participants in the final sample (Gender: 53 female, 48 male; Age: $M = 20.05$, $SD = 2.35$, Range = 17–35; Ethnicity: 6 Hispanic/Latino, 6 not Hispanic/Latino, 4 declined to answer; Race: 1 American Indian/Alaskan Native, 21 Asian, 7 Black/African American, 68 White, 4 other). No participants discontinued their participation in the experiment once the procedure had begun. Participants received either course credit or monetary compensation for their time. Informed consent was obtained from participants prior to each experiment, and this research was formally approved by the Purdue University Institutional Review Board.

2.2. Task

The reversal learning task was modeled after the study by Cools and colleagues (Cools, Altamirano, et al., 2006), and a conceptual overview is presented in Figure 1. Participants were instructed to imagine that they were the boss of a casino watching a customer playing a simple card game. The game uses only two cards, one depicting a face and one depicting a landscape scene. On each round, the dealer shuffles the cards and places them face down; the participant can see the cards, but the customer cannot. The customer chooses one card at random, which is highlighted on the screen by a gray border. Prior to each round, the dealer decides which of the two cards is the winning card and, depending on the dealer's decision and the customer's chosen card, the customer either wins \$100 or loses \$100; the dealer periodically changes their mind about which is the winning card. The participant's objective is to predict whether the customer will win or lose on that round, based on whether they think the customer has chosen the winning or losing card. The participant's response does not affect whether the customer wins or losses; their job is to simply observe and predict the outcomes. A key distinction here is that, unlike many other reversal learning paradigms, the participant does not receive reward or punishment feedback per se, but rather they receive feedback indicating whether the fictional customer has won or lost. That is, negative feedback on this task indicates that the customer lost on that trial; from the participant's perspective, this could represent correct feedback (if the rule remained unchanged and the participant accurately predicted the customer would lose) or error feedback (if the rule did change, and the participant incorrectly predicted the customer would win).

Images were drawn from the International Affective Picture System (#2200, 5201; Lang, Bradley, & Cuthbert, 2005). On each trial, the two images were presented vertically adjacent to one another with location randomized, indicating the two cards from the dealer. One image was highlighted by a gray border, indicating it was the customer's chosen card on that

trial (Figure 1). Participants were asked to predict whether the highlighted stimulus would lead to a reward or punishment for the customer on that trial by pressing either the left or right mouse button; images were presented until a behavioral response was made. A fixation mark ('+') was then presented for 1000 ms, followed by outcomes. Customer reward outcomes were depicted by a green smiley face and a +\$100 sign, while punishment outcomes were depicted by a red sad face and a -\$100 sign. Note that the outcomes depict whether the customer won or lost the round—not whether the participant correctly or incorrectly predicted the outcome. Outcome stimuli were presented for 500 ms and were followed by a fixation mark for 500 ms, after which the next trial began.

The task consisted of two conditions which differed in the valence of unexpected feedback: unexpected customer reward and unexpected punishment. The only difference between these conditions was the valence of the unexpected outcomes on reversal trials (i.e., the first trial in which the dealer changed the rule regarding the winning card). In the unexpected reward condition, all rule changes were signaled by unexpected positive outcomes, whereas in the unexpected punishment condition all rule changes were signaled by unexpected negative outcomes. Whereas positive and negative outcomes occurred in both conditions, *unexpected* positive outcomes occurred only in the reward condition, and *unexpected* negative outcomes occurred only in the punishment condition. Participants were not informed of this difference between conditions, and all other aspects of the task were the same across conditions. The order of conditions was randomized across participants. Here, we use the terms unexpected reward and unexpected punishment to refer to the two experimental conditions, even though the participant was not rewarded or punished per se.

Each condition consisted of an initial practice block (i.e., one separate practice block for each of the reward and punishment conditions) and 2 experimental blocks. Each practice block consisted of two stages: an initial acquisition stage (in which the rule was learned) and one reversal stage (in which the contingency changed). The learning criterion for the acquisition stage was 10 consecutive correct trials (i.e., the participant correctly predicted whether the customer would win or lose). Upon reaching this threshold, a *reversal trial* was presented, signaling to the participant that rule had changed (i.e., either unexpected reward or unexpected punishment, depending on the condition for that block). This was followed by a *switch trial*, in which participants were expected to adjust their behavioral response and predict the outcome based on the new rule (i.e., behavioral switch). Importantly, the same stimulus that had been highlighted on the reversal trial was highlighted again on the first behavioral switch trial; if participants made an incorrect prediction on the behavioral switch trial (indicating that they did not learn the new rule), the same stimulus was highlighted again on subsequent trials until a correct response was made. The learning criterion for this reversal stage was once again 10 consecutive correct trials, at which point participants proceeded to the experimental block. The practice block was terminated after 80 total trials if participants did not complete both the acquisition and reversal stages at that point.

Experimental blocks were different from practice blocks in two ways: the learning criteria and the number of stages. To make rule reversals less predictable, the learning criteria in the experimental blocks varied from 4–8 (as opposed to 10 in the practice blocks); these criteria were determined a priori according to a fixed pseudorandom sequence and were the same for

the two valence conditions for all subjects. Whereas the practice block contained only one initial acquisition stage and one reversal stage, experimental blocks continued until participants completed 100 trials. The number of stages completed in the experimental blocks, therefore, was dependent upon successful performance (i.e., participants adapting to the rule changes) and varied across participants. Participants completed 2 experimental blocks in each of the valence conditions (total reward trials: $M = 206.42$, $SD = 7.99$; total punishment trials; $M = 205.17$, $SD = 6.60$). Presentation software (Neurobehavioral Systems, Inc., Berkeley, CA) was used to control the timing and presentation of all stimuli.

2.3. Psychophysiological recording, data reduction, and analysis

The continuous EEG was recorded using an ActiCap and the ActiCHamp amplifier system (Brain Products GmbH, Munich, Germany). The EEG was digitized at 24-bit resolution and a sampling rate of 500 Hz. Recordings were taken from 32 scalp electrodes based on the 10/20 system, with a ground electrode at Fpz. Electrodes were referenced online to a virtual ground point (i.e., reference-free acquisition) formed within the amplifier (Luck, 2014), a method used in previous studies (Healy, Boran, & Smeaton, 2015; Novak & Foti, 2015; Weiler, Hassall, Krigolson, & Heath, 2015). The electrooculogram was recorded from two auxiliary electrodes placed 1 cm above and below the left eye, forming a bipolar channel. Electrode impedances were kept below 30 kOhms.

Brain Vision Analyzer (Brain Products) was used for offline analysis. Data were referenced to the mastoid average and bandpass filtered from 0.01–30 Hz using Butterworth zero phase filters. The signal was segmented from –200 to 800 relative to outcome onset. Correction for blinks and eye movements was performed using a regression method (Gratton, Coles, & Donchin, 1983). Individual channels were rejected trialwise using a semiautomated procedure, with artifacts defined as a step of 50 μV , a 200 μV change within 200-ms intervals, or a change of $<0.5 \mu V$ within 100-ms intervals. Additional artifacts were identified visually. ERPs were averaged separately for each trial type (unexpected positive outcomes, unexpected negative outcomes, expected positive outcomes, expected negative outcomes), and corrected relative to the 200-ms prestimulus baseline window.

ERPs were scored using time-window averages at representative electrodes: The RewP from 250–300 ms at Cz, the P3 from 300–450 ms at Pz, and the LPP from 450–800 ms at Pz. ERPs to unexpected outcomes (i.e., reversal outcomes signaling rule switches) were examined separately by condition (i.e., unexpected positive outcomes, unexpected negative outcomes). Responses to expected outcomes on non-switch trials were examined separately by valence, averaged across conditions (i.e., expected positive outcomes averaged across the reward and punishment conditions, expected negative outcomes averaged across the reward and punishment conditions). ERP effects were analyzed using 2 (Expectancy: expected vs. unexpected) $\times 2$ (Valence: positive vs. negative outcomes) repeated-measures ANOVAs. Behavioral data were analyzed by considering the error rates across Trial Type (switch, non-switch positive, non-switch negative) and Condition (reward, punishment). Associations between ERPs and error rate were examined using bivariate correlation, as well as multiple linear regression to test for the unique effect of each ERP when included as simultaneous

predictors. Statistical analysis was performed using IBM SPSS Statistics (Version 21.0, IBM, Armonk, NY).

3. Results

3.1. Behavioral Performance

The average number of stages completed was similar across conditions (Reward: $M = 19.33$, $SD = 2.84$; Punishment: $M = 19.05$, $SD = 3.15$; $t(100) = 1.16$, $p = .25$; $d = .12$). This indicates that the manipulation of unexpected feedback valence did not systematically influence learning on average. Error rates were examined across Trial Type (switch, non-switch positive outcome, non-switch negative outcome) and Condition (reward, punishment). The main effect of Trial Type was significant ($F(2,200) = 95.59$, $p < .001$), as well as the Condition $\times$ Trial Type interaction ($F(2,200) = 4.72$, $p < .05$); the main effect Condition was not significant ($F(1,100) = .76$, $p = .39$). As expected, participants made more errors on behavioral switch trials (i.e., trials immediately following unexpected feedback) than on non-switch trials (i.e., trials in which the rule remained unchanged), which was true for both the reward (switch vs. non-switch-positive: $t(100) = 9.369$, $p < .001$, $d = 1.20$; switch vs. non-switch negative: $t(100) = 8.195$, $p < .001$, $d = 1.01$) and punishment conditions (switch vs. non-switch-positive: $t(100) = 7.06$, $p < .001$, $d = .78$; switch vs. non-switch negative: $t(100) = 6.73$, $p < .001$, $d = .80$). The difference in error rate between the reward and punishment conditions did not reach significance for any of the three trial types (p 's $> .05$, d 's $< .20$), although there was a trend toward more errors on behavioral switch trials in the reward versus punishment conditions ($t(100) = 1.87$, $p = .06$).

3.2. Event-Related Potentials

ERP waveforms for each trial type are presented in Figure 2, and scalp distributions are presented in Figure 3. Of interest were effects of Expectancy (expected vs. unexpected outcomes), Valence (positive vs. negative outcomes), and their interaction.

3.2.1. Reward positivity—The RewP was maximal approximately 275 ms after outcome onset. RewP amplitude was sensitive to Valence ($F(1,100) = 5.87$, $p < .05$) but not Expectancy ($F(1,100) = 1.35$, $p = .25$); the interaction was not significant ($F(1,100) = 1.15$, $p = .28$). Consistent with the extant literature, the RewP was more positive to reward versus punishment outcomes overall, regardless of expectancy (Figure 3, top left).

3.2.2. P3—The P3 was maximal at approximately 400 ms after outcome onset. Unlike the RewP, P3 amplitude was sensitive to Expectancy ($F(1,100) = 54.58$, $p < .001$) but not Valence ($F(1,100) = .13$, $p = .72$); the interaction was not significant ($F(1,100) = 1.72$, $p = .19$). The P3 was more positive for unexpected versus expected outcomes, regardless of valence (Figure 3, top right).

3.2.3. Late positive potential—Following the P3, the LPP was sustained throughout the 450–800 ms window. Unlike either the RewP or P3, the ANOVA on LPP amplitude yielded main effects for both Expectancy ($F(1,100) = 179.26$, $p < .001$) and Valence ($F(1,100) = 19.43$, $p < .001$); the interaction was not significant ($F(1,100) = .42$, $p = .52$). The LPP was more

positive for unexpected outcomes (regardless of valence) and for punishment outcomes (regardless of expectancy). These effects were additive, such that the largest LPP was observed for unexpected punishment (i.e., trials in which participants predicted that the customer would win, but instead they lost). Both contrasts are presented in Figure 3 (bottom).

3.3. Links between Behavioral and Neural Measures

We considered bivariate correlations between ERPs to unexpected outcomes (i.e., reversal trials in which the rule changed) and error rate on behavioral switch trials (i.e., trials immediately following reversal trials, in which participants were required to change their response) using Bonferroni adjusted alpha levels of .017 per test (.05/3). These correlations were calculated separately for the reward and punishment conditions to test for specific effects of unexpected outcome valence. Scatter plots depicting these bivariate correlations are presented in Figure 4. In the reward condition, higher error rate was associated with reduced ERP amplitudes across all three components (RewP: $r = -.29$, $p < .01$; P3: $r = -.48$, $p < .001$; LPP: $r = -.39$, $p < .001$). A similar pattern was found for the punishment condition (RewP: $r = -.19$, $p = .06$; P3: $r = -.39$, $p < .001$; LPP: $r = -.33$, $p < .001$).¹ This indicates that a blunted overall ERP response to unexpected outcomes predicted the likelihood of committing errors on behavioral switch trials (i.e., failure to adapt to the new rule).

In light of these correlations across all three components, we then examined potential unique associations using multiple regression. First, we tested for unique effects of each ERP component *within* the reward and punishment conditions by including the RewP, P3, and LPP in that condition (unexpected trials) as simultaneous predictors of error rate. In the reward condition, a significant effect was found for the P3 ($\beta = -.52$, $p < .05$) but not the RewP ($\beta = .12$, $p = .35$) or LPP ($\beta = -.06$, $p = .68$). The identical pattern was found in the punishment condition, with a significant effect only for the P3 ($\beta = -.41$, $p < .05$; RewP: $\beta = .12$, $p = .33$; LPP: $\beta = -.09$, $p = .53$).² This indicates that within each condition, the strongest association with behavioral performance was found with P3 amplitude.

We further explored the link between P3 amplitude and performance by testing for unique effects of *valence* for this ERP. Specifically, we included the P3 to unexpected reward and the P3 to unexpected punishment as simultaneous predictors of behavioral switch trial error rate, calculated separately for the reward and punishment conditions. Predicting error rate in the reward condition, a unique effect was observed only for the P3 to unexpected reward ($\beta = -.56$, $p < .001$) and not the P3 to unexpected punishment ($\beta = .13$, $p = .24$). Conversely, predicting error rate in the punishment condition, a unique effect was observed only for the P3 to unexpected punishment ($\beta = -.30$, $p < .05$) and not the P3 to unexpected reward ($\beta = -.15$, $p = .20$). This indicates that, while the P3 was insensitive to valence at the mean level

¹We also examined relationships between ERPs and reaction time (i.e., switch trials) while adjusting for multiple comparisons (using Bonferroni adjusted alpha level of .017); however, no associations were significant in the reward condition (RewP: $r = .10$, $p = .31$; P3: $r = .12$, $p = .22$; LPP: $r = .15$, $p = .14$). In the punishment condition, enhanced P3 amplitude to unexpected outcomes was significantly related to slower reaction times ($r = .24$, $p = .016$); no other ERPs were significant (RewP: $r = .20$, $p = .043$; LPP: $r = .19$, $p = .052$).

²We did not find unique effects of ERPs predicting reaction time (i.e., switch trials) in the reward condition (RewP: $\beta = .03$, $p = .86$; P3: $\beta = .003$, $p = .99$; LPP: $\beta = .13$, $p = .40$). Furthermore, No significant unique effects of ERPs predicting reaction time on switch trials in the punishment condition were found (RewP: $\beta = .08$, $p = .53$; P3: $\beta = .17$, $p = .31$; LPP: $\beta = .03$, $p = .86$).

(i.e., averaged across subjects), there were valence-specific associations between P3 amplitude and behavioral adjustment on switch trials following unexpected outcomes. Reward learning is best predicted by the P3 to unexpected rewards (but not unexpected punishment), and punishment learning is best predicted by the P3 to unexpected punishment (but not reward).

4. Discussion

The current study sheds new light on the neural correlates of reversal learning in three key ways: (1) Differential patterns of feedback valence and expectancy were observed across stages of processing. RewP amplitude was sensitive to feedback valence but not expectancy, P3 amplitude was sensitive to expectancy but not valence, and LPP amplitude was sensitive to both (but not their interaction). This indicates that all three ERPs are differentially sensitive to the characteristics of environmental feedback. (2) Links between ERPs and behavioral performance were strongest for the P3: a blunted P3 to unexpected feedback predicted a higher error rate on subsequent behavioral switch trials, even after controlling for RewP and LPP amplitudes. (3) The association between P3 amplitude and behavioral performance was valence-specific, such that error rate in the reward condition was specifically predicted by the P3 to unexpected reward, and the error rate in the punishment condition was specifically predicted by the P3 to unexpected punishment.

This study represents the first attempt at a simultaneous examination of the RewP, P3, and LPP using a deterministic reversal-learning paradigm capable of isolating valence-specific effects. This enabled us to link distinct stages of processing to behavioral performance. This serves as a significant follow-up to Chase and colleagues (2011) who found significant links between ERP amplitude and behavioral performance; however, their paradigm consisted of only a negative feedback condition. Therefore, they were unable to properly capture valence-specific effects. To the best of our knowledge, this is the first time an endogenous ERP component such as the P3 has been linked to behavioral adjustment following contingency reversals *and* identifies unique valence effects.

In contrast with several previous studies, we also disentangled two ERP components (i.e., P3, LPP) whose relationship to behavioral responses is either unknown (i.e., did not examine LPP; Chase et al., 2011) or has been made ambiguous (San Martín et al., 2013). Specifically, by scoring the P3 at a broad time interval (e.g., 400–800ms), it is likely that some previous studies unintentionally captured two distinct stages of processing (i.e., P3 and LPP), thus obscuring their findings (San Martín et al., 2013). By teasing apart the P3 (300–450ms at Pz) and LPP (450–800ms at Pz) as discreet components, we provide evidence of the unique effect of the P3 and novel evidence for the characteristics of the LPP during reversal-based learning. The P3 in this study reflects a quantitative indicator of attention allocation devoted to feedback stimuli. Thus, from a psychological standpoint lapses in attention (reduced P3) toward unexpected feedback (reward and punishment) may explain poorer behavioral adjustment on subsequent learning trials.

Participants were instructed to observe and respond to feedback as an objective third person (i.e., boss of a casino); thus, given these parameters one might expect that the RewP was

tracking feedback expectancy (i.e., correct/incorrect) rather than valence (i.e., reward or punishment). For example, expected feedback could be viewed as ‘positive’ from the participant’s perspective because it affirms their expectations, regardless of whether it is a reward or loss in this context. Conversely, unexpected feedback could be viewed as ‘negative’. Recently however, one study postulated that different brain systems each receive their own independent feedback signal on every trial (Helie, Paul, & Ashbey, 2011), namely, the prefrontal cortex and basal ganglia. Under this conceptual model, self-generated second-order feedback are likely processed in the prefrontal cortex (e.g., I expected a reward, the outcome is a reward, therefore my response was correct), whereas the basal ganglia is thought reflect outcome valence (e.g., the outcome was good). RewP amplitude characteristics seem to be more consistent with the latter system, and a number of studies implicate the striatum as a possible neural generator (Carlson, Foti, Mujica-Parodi, Harmon-Jones, & Hajcak, 2011; Foti et al., 2011; Becker, Nitsch, Miltner, & Straube, 2014). Consistent with these RewP interpretations, an observational ERP study found similar RewP amplitude activity between observers and performers, even when there is no consequent of the performer’s behavior for the observer (Marco-Pallarés, Krämer, Strehl, Schröder, & Münte, 2010); thus demonstrating the RewP is capable of tracking outcome valence regardless even if they are not receive feedback first had. This is inline with the findings in the present study, which found effects of valence with respect to the RewP.

One of the implications of these findings is the potential clinical utility of the unique valence P3 effects to unexpected feedback. This reversal learning paradigm in clinical populations was used with people affected by Parkinson’s disease (PD; Cools et al. 2006); however, no study has explored time-course of reversal learning among this group. It would be of interest to examine whether the P3 of PD patients is affected by the valence of the stimulus, given that some have reported PD patients were less sensitive to unexpected punishment than to unexpected reward (Cools et al. 2006). In addition to PD, it is possible that deficits in reversal learning may also be linked to other psychiatric disorders characterized by dopaminergic dysfunction, including major depression (Remijnse et al., 2009; Robinson et al., 2012) and pathological gambling (Hewig et al., 2010). Indeed, hyposensitivity to rewards has been implicated in depression (Pizzagalli, 2013; Treadway & Zald, 2011) and a wealth of evidence indicates that depression is characterized by a diminished P3 (Bruder, Kayser, & Tenke, 2012). Conversely, pathological gambling has been linked not only to hypersensitivity to rewards (Hewig et al., 2010), but also impaired reversal learning (Patterson, Holland, & Middleton, 2006). It will be important for future research to parse apart the unique association between the P3 and behavioral performance across these clinical dimensions. It is possible that a reduced P3 to unexpected rewards serves as a neural indicator for poor behavioral performance among those with depression, whereas reduced P3 to unexpected punishment is characteristic of pathological gamblers; the current paradigm is well-designed to test this possibility, and further analyses of individual differences in ERPs and performance on this task are underway in a larger sample.

Another implication of the findings concerns reinforcement learning (RL). RL is a feedback-driven (instrumental) type of learning commonly referred to as trial-and-error learning. RL has been the focus of much work in cognitive neuroscience, and one of the most well-established roles of dopamine (DA) in cognitive function is related to feedback in RL

(Ashby & Helie, 2011; Schultz, Dayan, & Montague, 1997). Specifically, DA levels in the basal ganglia increase when an unexpected reward is received, dip when an expected reward fails to appear, and change with a high-enough temporal resolution to serve as a reward signal in RL (Helie, Ell, & Ashby, 2015). In the work included in this article, the RewP was sensitive to feedback valence and could reflect the role of DA in the ventral striatum. Helie et al. (2012) argued using a computational model that the striatum processes the feedback valence in this task, and there is evidence suggesting that RewP could be generated by the basal ganglia (Foti et al., 2011). In contrast, the P3 may reflect a more processed (cortical) type of feedback (Helie et al., 2012). Hence, the RewP may be useful as a proxy measure of DA in instrumental learning tasks (which would be difficult to measure in real-time in humans) and facilitate studies of reinforcement learning in humans. More work is needed to validate the use of the RewP as a proxy to measure DA in RL tasks.

Our findings also highlight an interesting avenue for future studies of utilizing the P3 in relation to temporal difference (TD) learning (Montague, Dayan, & Sejnowski, 1996), a method that has been used to address the RL problem. TD theory is linked to concepts pertaining to incentive salience, which is the process by which stimuli grasps attention and motivates goal directed behavior (Berridge, 2007). Put simply, TD learning predicts that participants learn to anticipate rewards, and that phasic DA spikes can be detected before the actual reward is presented. However, TD learning relies on the assumption that the cue allowing for anticipation of the reward is noticed. It is possible that if P3 was used in conjunction with studies such as Kumar et al (2008), we may find reduced P3 amplitude (i.e., failure to reallocate attention to salient stimuli; Courchesne et al., 1975) to be associated with reduced phasic TD reward learning signals, particularly in non-brainstem regions of depressed populations; however, future research is needed to test this possibility.

A strength of the present study is the large sample size ($N=101$), which to the best of our knowledge is among the largest reversal learning ERP study to date. The relatively large sample size was strength to detect correlations between ERP amplitudes and behavioral performance on this reversal learning paradigm. It was unclear from previous research what the strength of the effects were, as such it was important to be adequately powered to detect relatively small effects, if they were present. Previous studies in this area of research have been characterized by relatively small sample sizes ($N=13$; Chase et al., 2011), which greatly limits the ability to detect individual differences in performance and ERP amplitudes. It is important that the methodological strengths of the present study be viewed in light of the following limitation: the present study utilized a relatively healthy undergraduate and community sample. It will be important to replicate these findings in individuals with varying presenting psychopathologies to determine whether the effects generalize to or differ across groups. While it was important to examine this paradigm in healthy populations, separate studies need to be conducted to determine whether this task is valid in other groups (e.g., depression, pathological gambling).

5. Conclusions

Deficits in reversal learning are thought to play an important role in adapting one's responses to changing environmental contingencies. The current study extends this literature

by examining the temporal dynamics of reversal-based learning using both behavioral and neural indicators. We found evidence of distinct patterns of environmental feedback valence and expectancy across three stages of processing (i.e., RewP, P3, LPP). Furthermore, we found that the P3 was most strongly related to behavioral performance, such that a diminished P3 to unexpected feedback predicted higher error rates on future behavioral switch trials. Finally, we found distinct effects of valence between behavioral response and P3 amplitude: the P3 to unexpected reward specifically predicts error rate in reward conditions, while the P3 to unexpected punishment specifically predicts error rate in punishment conditions. Overall, P3 is likely a reliable neural candidate of examining error processing in reversal-based learning across multiple clinical dimensions.

Acknowledgments

This research was funded, in part, by NIH Grant 2R01MH063760-09A1 to Sebastien Hélie.

References

- Aron AR, Poldrack RA. Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *The Journal of Neuroscience*. 2006; 26(9):2424–2433. [PubMed: 16510720]
- Ashby FG, Helie S. A tutorial on computational cognitive neuroscience: Modeling the neurodynamics of cognition. *Journal of Mathematical Psychology*. 2011; 55:273–289. [PubMed: 21841845]
- Becker MP, Nitsch AM, Miltner WH, Straube T. A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. *The Journal of Neuroscience*. 2014; 34(8):3005–3012. [PubMed: 24553940]
- Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*. 2007; 191(3):391–431. [PubMed: 17072591]
- Bruder GE, Kayser J, Tenke CE. Event-related brain potentials in depression: Clinical, cognitive and neurophysiologic implications. *The Oxford Handbook of Event-Related Potential Components*. 2012; 2012:563–592.
- Carlson JM, Foti D, Mujica-Parodi LR, Harmon-Jones E, Hajcak G. Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. *Neuroimage*. 2011; 57(4):1608–1616. [PubMed: 21624476]
- Chase HW, Swainson R, Durham L, Benham L, Cools R. Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience*. 2011; 23(4):936–946. <http://doi.org/10.1162/jocn.2010.21456>. [PubMed: 20146610]
- Codispoti M, Ferrari V, Bradley MM. Repetition and event-related potentials: distinguishing early and late processes in affective picture perception. *Journal of Cognitive Neuroscience*. 2007; 19(4):577–586. <http://doi.org/10.1162/jocn.2007.19.4.577>. [PubMed: 17381249]
- Cools R, Altamirano L, D'Esposito M. Reversal learning in Parkinson's disease depends on medication status and outcome valence. *Neuropsychologia*. 2006; 44(10):1663–1673. <http://doi.org/10.1016/j.neuropsychologia.2006.03.030>. [PubMed: 16730032]
- Cools R, Clark L, Owen AM, Robbins TW. Defining the Neural Mechanisms of Probabilistic Reversal Learning Using Event-Related Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*. 2002; 22(11):4563–4567. [PubMed: 12040063]
- Cools R, Lewis SJG, Clark L, Barker RA, Robbins TW. L-DOPA Disrupts Activity in the Nucleus Accumbens during Reversal Learning in Parkinson's Disease. *Neuropsychopharmacology*. 2006; 32(1):180–189. <http://doi.org/10.1038/sj.npp.1301153>. [PubMed: 16841074]
- Courchesne E, Hillyard SA, Galambos R. Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*. 1975; 39(2):131–143. [PubMed: 50210]

- Donchin E, Coles MGH. Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*. 1988; 11(03):357–374. <http://doi.org/10.1017/S0140525X00058027>.
- Foti D, Hajcak G, Dien J. Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. *Psychophysiology*. 2009; 46(3):521–530. [PubMed: 19496228]
- Foti D, Weinberg A, Dien J, Hajcak G. Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospacial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*. 2011; 32(12):2207–2216. <http://doi.org/10.1002/hbm.21182>. [PubMed: 21305664]
- Gehring WJ, Willoughby AR. The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses. *Science*. 2002; 295(5563):2279–2282. <http://doi.org/10.1126/science.1066893>. [PubMed: 11910116]
- Gratton G, Coles MG, Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*. 1983; 55(4):468–484. [PubMed: 6187540]
- Hajcak G, Dunning JP, Foti D. Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology*. 2009; 120(3):505–510. <http://doi.org/10.1016/j.clinph.2008.11.028>. [PubMed: 19157974]
- Hajcak G, Olvet DM. The persistence of attention to emotion: brain potentials during and after picture presentation. *Emotion (Washington, DC)*. 2008; 8(2):250–255. <http://doi.org/10.1037/1528-3542.8.2.250>.
- Healy GF, Boran L, Smeaton AF. Neural patterns of the implicit association test. *Frontiers in human neuroscience*. 2015; 9
- Helie S, Ell SW, Ashby FG. Learning robust cortico-frontal associations with the basal ganglia: An integrative review. *Cortex*. 2015; 64:123–135. [PubMed: 25461713]
- Helie S, Paul EJ, Ashby FG. Simulating the effect of dopamine imbalance on cognition: From positive affect to Parkinson's disease. *Neural Networks*. 2012; 32:74–85. [PubMed: 22402326]
- Hewig J, Kretschmer N, Trippe RH, Hecht H, Coles MG, Holroyd CB, Miltner WH. Hypersensitivity to reward in problem gamblers. *Biological Psychiatry*. 2010; 67(8):781–783. [PubMed: 20044073]
- Holroyd CB, Pakzad-Vaezi KL, Krigolson OE. The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*. 2008; 45(5): 688–697. <http://doi.org/10.1111/j.1469-8986.2008.00668.x>. [PubMed: 18513364]
- Lang, P., Bradley, M., Cuthbert, B. International affective picture system (IAPS): affective ratings of pictures and instruction manual (tech report A-6). University of Florida. Center for Research in Psychophysiology; Gainesville: 2005.
- Luck, SJ. An introduction to the event-related potential technique. MIT press; 2014.
- Marco-Pallarés J, Krämer UM, Strehl S, Schröder A, Münte TF. When decisions of others matter to me: an electrophysiological analysis. *BMC neuroscience*. 2010; 11(1):86. [PubMed: 20670398]
- Miltner WH, Braun CH, Coles MG. 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*. 1997; 9(6):788–798. <http://doi.org/10.1162/jocn.1997.9.6.788>. [PubMed: 23964600]
- Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of neuroscience*. 1996; 16(5):1936–1947. [PubMed: 8774460]
- Novak KD, Foti D. Teasing apart the anticipatory and consummatory processing of monetary incentives: An event-related potential study of reward dynamics. *Psychophysiology*. 2015; 52(11): 1470–1482. [PubMed: 26223291]
- Patterson J 2nd, Holland J, Middleton R. Neuropsychological performance, impulsivity, and comorbid psychiatric illness in patients with pathological gambling undergoing treatment at the CORE Inpatient Treatment Center. *Southern Medical Journal*. 2006; 99(1):36–43. [PubMed: 16466120]
- Patzelt EH, Kurth-Nelson Z, Lim KO, MacDonald AW. Excessive state switching underlies reversal learning deficits in cocaine users. *Drug and Alcohol Dependence*. 2014; 134 <http://doi.org/10.1016/j.drugalcdep.2013.09.029>.
- Proudfit GH. The reward positivity: from basic research on reward to a biomarker for depression. *Psychophysiology*. 2015; 52(4):449–459. <http://doi.org/10.1111/psyp.12370>. [PubMed: 25327938]

- Remijnse PL, Nielen MMA, van Balkom AJLM, Hendriks GJ, Hoogendijk WJ, Uylings HBM, Veltman DJ. Differential frontal–striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive–compulsive disorder. *Psychological Medicine*. 2009; 39(09): 1503–1518. <http://doi.org/10.1017/S0033291708005072>. [PubMed: 19171077]
- Robinson OJ, Cools R, Carlisi CO, Sahakian BJ, Drevets WC. Ventral Striatum Response During Reward and Punishment Reversal Learning in Unmedicated Major Depressive Disorder. *American Journal of Psychiatry*. 2012; 169(2):152–159. <http://doi.org/10.1176/appi.ajp.2011.11010137>. [PubMed: 22420038]
- Robinson OJ, Frank MJ, Sahakian BJ, Cools R. Dissociable responses to punishment in distinct striatal regions during reversal learning. *NeuroImage*. 2010; 51(4):1459–1467. <http://doi.org/10.1016/j.neuroimage.2010.03.036>. [PubMed: 20303408]
- San Martín R, Appelbaum LG, Pearson JM, Huettel SA, Woldorff MG. Rapid brain responses independently predict gain maximization and loss minimization during economic decision making. *The Journal of Neuroscience*. 2013; 33(16):7011–7019. [PubMed: 23595758]
- Schultz W, Dayan P, Montague RRA. Neural substrate of prediction and reward. *Science*. 1997; 275:1593–1599. [PubMed: 9054347]
- Schultz W, Tremblay L, Hollerman JR. Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex*. 2000; 10(3):272–283. [PubMed: 10731222]
- Torres A, Catena A, Cándido A, Maldonado A, Megías A, Perales JC. Cocaine Dependent Individuals and Gamblers Present Different Associative Learning Anomalies in Feedback-Driven Decision Making: A Behavioral and ERP Study. *Frontiers in Psychology*. 2013; 4:122. <http://doi.org/10.3389/fpsyg.2013.00122>. [PubMed: 23516173]
- von Borries AKL, Verkes RJ, Bulten BH, Cools R, de Bruijn ERA. Feedback-related negativity codes outcome valence, but not outcome expectancy, during reversal learning. *Cognitive, Affective & Behavioral Neuroscience*. 2013; 13(4):737–746. <http://doi.org/10.3758/s13415-013-0150-1>.
- Wallis JD, Kennerley SW. Heterogeneous reward signals in prefrontal cortex. *Current opinion in neurobiology*. 2010; 20(2):191–198. [PubMed: 20303739]
- Weiler J, Hassall CD, Krigolson OE, Heath M. The unidirectional prosaccade switch-cost: Electroencephalographic evidence of task-set inertia in oculomotor control. *Behavioural brain research*. 2015; 278:323–329. [PubMed: 25453741]
- Weinberg A, Hajcak G. Beyond good and evil: the time-course of neural activity elicited by specific picture content. *Emotion (Washington, DC)*. 2010; 10(6):767–782. <http://doi.org/10.1037/a0020242>.

Figure 1.

Sample task trial. Participants ($N = 101$) were presented with two images, one of which was highlighted by a gray box (here, the landscape). The highlighted stimulus indicated which was chosen by the customer on that trial, and the participant was instructed to predict whether the customer would win or lose based on their choice. Participants clicked either the left or right mouse button according to whether they predicted reward or punishment for the customer. Following this response, participants were presented with the customer's outcome (here, reward).

Figure 2.

ERP waveforms presented separately by condition at electrodes Cz (top) and Pz (bottom) for the full sample ($N = 101$). Expected outcomes indicate behavioral non-switch trials in which the rule remained unchanged. Unexpected outcomes indicate reversal trials in which the rule changed unexpectedly. Reward and punishment refer to the valence of the presented outcome, indicating that the fictional customer (not the participant) won or lost money on that trial.

Figure 3. Scalp distributions representing the difference between the relevant conditions for each ERP across the full sample ($N = 101$), based on the omnibus ANOVAs. Note that the scales are different for each headmap.

Figure 4. Scatterplots depicting the bivariate correlations between ERPs (i.e., RewP [top], P3 [middle], LPP [bottom]) to unexpected outcomes (i.e., reversal trials in which the rule changed) and error rate on behavioral switch trials for the reward (left) and punishment (right) condition for the full sample ($N=101$). Correlations within each condition (i.e., reward, punishment) were calculated separately using Bonferroni adjusted alpha levels of .017 per test (.05/3).