Mediofrontal Negativity Signals Unexpected Timing of Salient Outcomes

Sara Garofalo[1,2], Christopher Timmermann[1,3], Simone Battaglia[1], Martin E. Maier[1,4], and Giuseppe di Pellegrino[1,5]

Abstract

■ The medial prefrontal cortex (mPFC) and ACC have been consistently implicated in learning predictions of future outcomes and signaling prediction errors (i.e., unexpected deviations from such predictions). A computational model of ACC/mPFC posits that these prediction errors should be modulated by outcomes occurring at unexpected times, even if the outcomes themselves are predicted. However, unexpectedness per se is not the only variable that modulates ACC/mPFC activity, as studies reported its sensitivity to the salience of outcomes. In this study, mediofrontal negativity, a component of the event-related brain potential generated in ACC/mPFC and coding for prediction errors, was measured in 48 participants performing a Pavlovian aversive conditioning task, during which aversive (thus salient) and neutral outcomes were unexpectedly shifted (i.e., anticipated or delayed) in time. Mediofrontal ERP signals of prediction error were observed for outcomes occurring at unexpected times but were specific for salient (shock-associated), as compared with neutral, outcomes. These findings have important implications for the theoretical accounts of ACC/mPFC and suggest a critical role of timing and salience information in prediction error signaling. ■

INTRODUCTION

Accurate prediction of future outcomes is crucial to achieve optimal behavior in an uncertain and dynamic environment. Modern theories of conditioning (Pearce & Hall, 1980; Mackintosh, 1975; Rescorla & Wagner, 1972), computational explanation (Sutton & Barto, 1998), and neural data (Schultz & Dickinson, 2000) suggest that the process of formation and evaluation of predictions is driven by learning signals, such as prediction errors, which signal errors in the prediction of outcomes and thus the need for updating these predictions.

Medial prefrontal cortex (mPFC) has been repeatedly and extensively implicated in representing and updating the value of outcomes (Silvetti, Seurinck, & Verguts, 2013; Rushworth & Behrens, 2008).

A computational model of the mPFC, the predicted response–outcome (PRO) model (Alexander & Brown, 2011, 2014), contends that mPFC, and especially ACC, is critically involved in learning predictions of future outcomes and signaling unexpected deviations from such predictions (i.e., prediction errors), irrespective of their valence. Specifically, the model suggests that some ACC/mPFC neurons are reliably activated as a consequence of specific events and increase their firing to anticipate them, peaking around the time when the event is most likely to occur. The more likely an event becomes, the stronger the anticipatory neural firing also becomes. These signals reflect a learned prediction of possible outcomes and are inhibited when the outcome actually occurs. Consequently, maximal activity will be registered when an expected outcome fails to occur, irrespective of whether the outcome is bad or good. Cells that are excited by predictions and inhibited by outcomes will generate prediction error signals, computed as the difference between prediction and outcome.

Interestingly, the PRO model proposes that predictions concern not only the likelihood of occurrence of outcomes but also their timing. Thus, ACC/mPFC activation should be registered for outcomes occurring at unexpected times, even if the outcomes themselves are expected (Alexander & Brown, 2011), therefore signaling a timing prediction error. Predictions about the timing of outcome delivery are the result of implicit learning processes in which relevance is given to not only which event will occur or how likely its occurrence is but also when it will occur. In this context, prediction error signals are used to update previous expectations, when new contradictory and salient events occur.

This assumption of the PRO model appears largely coherent with single-unit studies in monkey mPFC that report precisely timed patterns of activation before the occurrence of an outcome (Shidara & Richmond, 2002; Niki & Watanabe, 1979) and further increase of firing whenthe expected outcome is omitted (Amador, Schlag-Rey, & Schlag, 2000).

[1]University of Bologna, [2]University of Cambridge, [3]Imperial College London, [4]Catholic University of Eichstätt-Ingolstadt, [5]Bangor University

In line with single-unit data, fMRI studies in humans revealed that ventral tegmental area (VTA) activity codes for time-dependent reward prediction error (Klein-Flügge, Hunt, Bach, Dolan, & Behrens, 2011) and that regions within the mPFC are responsive to unexpectedly delayed outcomes (Forster & Brown, 2011).

Critical for the present purposes, electrophysiological correlates of prediction error signals can also be identified in negative ERPs recorded frontocentrally and peaking between 200 and 350 msec after unexpected events, such as the feedback-related negativity (FRN; Sambrook & Goslin, 2016; Hauser et al., 2014; for reviews, see Walsh & Anderson, 2012; Nieuwenhuis, Holroyd, Mol, & Coles, 2004).

Several studies on mediofrontal negativity have provided empirical evidence supporting the PRO model proposal that ACC/mPFC plays a general role in predicting the likelihood of occurrence of outcomes and signaling prediction errors, irrespective of their valence (Garofalo, Maier, & di Pellegrino, 2014; Talmi, Atkinson, & El-Deredy, 2013; Ferdinand, Mecklinger, Kray, & Gehring, 2012; Jessup, Busemeyer, & Brown, 2010; Oliveira, McDonald, & Goodman, 2007). By comparison, little is known about mediofrontal ERP sensitivity to unexpected variations of the time between predictive event and resulting outcome.

However, unexpectedness per se is not the only variable that accounts for ACC/mPFC activation. Mediofrontal negative ERP components have been reported to be sensitive to the salience of outcomes and errors, being modulated by the amount of punishment (Maier & Steinhauser, 2013; Talmi et al., 2013; Ganushchak & Schiller, 2008; Hajcak, Moser, Yeung, & Simons, 2005), the attentional significance (Maier, di Pellegrino, & Steinhauser, 2012; Maier, Yeung, & Steinhauser, 2011), or the perceptual salience (Lou, Hsu, & Sajda, 2015) of the events regardless of their likelihood.

Accordingly, the aim of this study is to test whether unexpected timing of salient (i.e., aversive) outcomes, as compared with neutral outcomes, can trigger ACC/mPFC activity expressed as mediofrontal negativity. To this aim, EEG activity was recorded from 48 participants performing a Pavlovian aversive conditioning task. Visual conditioned stimuli (CS) were associated with aversive (i.e., shock-associated) or neutral outcomes and delivered at different timings: In 80% of the trials, the outcome was presented at a predicted timing; in 20% of the trials, the outcome was unexpectedly shifted in time (either anticipated or delayed).

If only unexpectedness of events accounts for ACC/mPFC activity (as proposed by the PRO model), then salient and neutral stimuli delivered at unexpected times should be equally able to elicit mediofrontal negativity. Whereas, if unexpected and additionally salient outcomes selectively trigger ACC/mPFC activity, these should be able to elicit a stronger mediofrontal negativity, as compared with neutral outcomes, when unexpectedly shifted in time.

METHODS

Participants

Forty-eight volunteers (24 women; 8 left-handed; mean age = 23.62 years, *SD* = 2.39 years) were recruited from the student population of the University of Bologna and enrolled for the study. The participants had no history of neurological diseases and normal or corrected-to-normal vision. All gave written informed consent before the beginning of the experiment. The study was conducted in accordance with institutional guidelines of the University of Bologna and was approved by the ethics committee of the Department of Psychology.

Stimuli and Materials

A Pavlovian aversive conditioning paradigm was used for the task. Japanese Kanji symbols served as conditioned (CS+) and neutral (CS−) stimuli. Mild electric shocks were used as unconditioned stimuli (US). Colored frames surrounding the CSs served as visual outcomes signaling to the participants the upcoming arrival of outcome. The visual outcomes were of two possible colors (blue or orange), and each color was uniquely associated with either aversive (CS+), thus shock-associated or neutral (CS−) outcome. Colors were counterbalanced across participants. The stimuli were displayed at a viewing distance of 80 cm on a 17-in. color monitor with a black background. A personal computer running Presentation software (Neurobehavioral Systems, Albany, CA) controlled stimulus presentation. All stimuli were balanced for luminance, complexity, and color saturation. A Digitimer Stimulator (Model DS7; Digitimer Ltd., Hertfordshire, UK) generated the electrical shocks, consisting of pulses of 200 msec. Two SU15N1 electrodes (SEI EMG, Padova, Italy), attached to the participants' left hand, were used for delivering mild shocks. Shock intensity was adjusted for each participant before each block: Stimulation intensity was initially set at 0.5 and gradually increased (1-mA increments) to a level perceived as "annoying, but not painful" by the participant. The mean shock intensity was 3.58 mA (*SD* = 3.08 mA, min = 0.32 mA, max = 9.27 mA).

Task Description

Two versions of the task were created, differing solely in the timing of the visual outcomes: an early-outcome task and a late-outcome task (Figure 1). Participants had to complete only one version, and possible differences were tested with a between-participant design.

The task began displaying a black screen for a random interval (1–2 sec), followed by a fixation cross. Participants were required to press a button when presented with the fixation cross to start the trial. For each trial, a stimulus (CS+/CS−) was displayed for a variable amount of time (depending on the early-/late-outcome task) and then followed by the corresponding visual outcome,

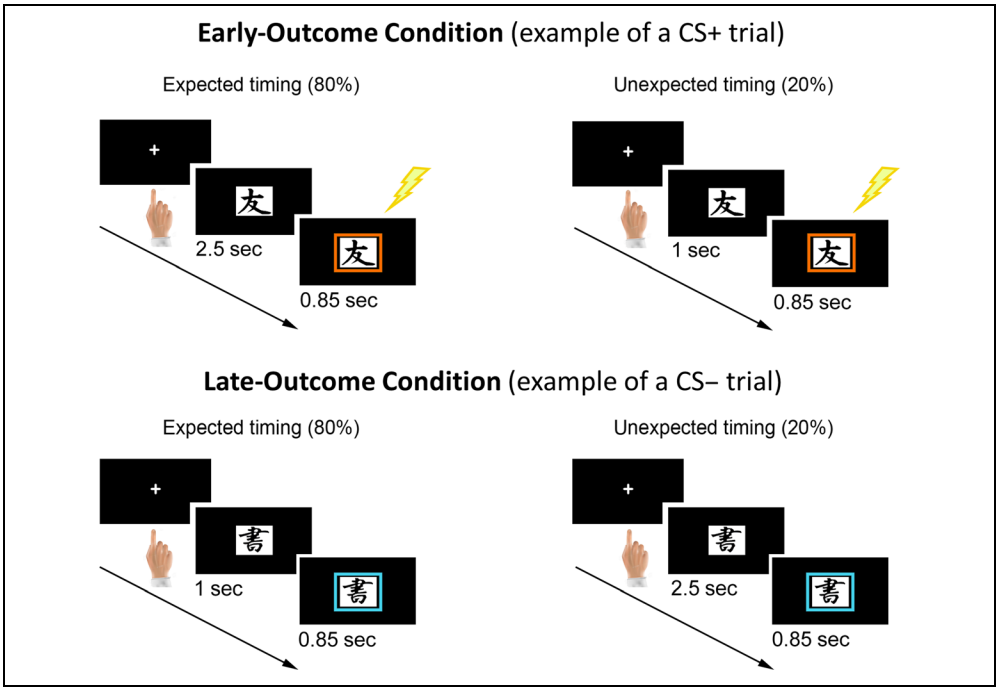

Figure 1. Pavlovian aversive conditioning task. Participants were required to press a button to start a new trial and were subsequently presented with one of two possible CSs (CS+/CS−). The stimuli were followed by a visual outcome indicating an imminent shock delivery (shock trials: CS+) or neutral outcome (neutral trials: CS−). Such outcomes occurred on 80% of the trials at an expected timing (left) and were shifted in time on 20% of the trials (right). Two task versions were used between participants: early (top) and late (bottom) outcome conditions, differing solely on the timing in which outcomes were presented. Note that, although the figure depicts CS+ trials in the early-outcome condition and CS− trials in late-outcome condition, both types of trials were present in each version of the task.

consisting of a colored frame (850 msec). During shock trials (CS+), a shock (US) was delivered in the last 200 msec of the visual outcome. During neutral trials (CS−), the visual outcome was not paired with anything. As the mediofrontal negativity occurs within about 400 msec after outcome onset, an interval of 600 msec between the onset of the visual outcome and US delivery allowed to avoid confounding effects because of the overlap between shock delivery and the epoch of interest for the EEG analysis.

In the early-outcome task, the CS was followed by the visual outcome after 2500 msec in 80% of the trials (expected timing) and after 1000 msec in 20% of the trials (unexpected timing). Consequently, in a lower number of trials, the outcome was unexpectedly anticipated in time. In the late-outcome task, the CS was followed by the corresponding outcome after 1000 msec in 80% of the trials (expected timing) and after 2500 msec in 20% of the trials (unexpected timing). Consequently, in a lower number of trials, the outcome was unexpectedly delayed in time.

The task consisted of 858 trials, divided in 11 blocks of 78 trials each. Each block included half CS+ trials and half CS− trials. Each block began with a learning phase, during which only the expected timing condition was presented for the first 10 trials. In 60 trials (of the 68 remaining trials), the CSs were paired with an outcome, and in eight trials, the CSs were not paired with an outcome, resulting in a catch trial proportion of approximately 12%. The 60 trials with an outcome followed the 80–20 timing schedule explained earlier (48 trials with expected timing and 12 trials with unexpected timing). The trials in the learning phase were not considered for the EEG analysis. Skin conductance response (SCR) was analyzed on catch trials only.

To assess explicit learning, at the end of each block, participants were required to pair each stimulus presented in that block with its own outcome.

Procedure

Participants were comfortably seated in a silent room, and their position was centered relative to the screen. EEG and SCR were recorded continuously while participants completed the task and data were stored for offline analysis. Participants were asked to remain as quiet and still as possible during task completion and keep their gaze at the center of the screen. After verifying that EEG and SCR were being properly recorded, the intensity of shock delivery was adjusted for each participant (see Stimuli and Materials). Written instructions for completing the task appeared on the computer screen. Participants were informed that their responses and actions had no effect on shock administration and were allowed to take short breaks between blocks.

EEG Recording and Analysis

The EEG signal was recorded using Ag–AgCl electrodes (Fast'n Easy Electrodes; Brain Products, Gilching, Germany) at 59 electrode sites (Fp1, Fp2, AF3, AF4, AF7, AF8, F1, F2, F3, F4, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FT7, FT8, C1, C2, C3, C4, C5, C6, T7, T8, CP1, CP2, CP3, CP4, CP5, CP6,

TP7, TP8, P1, P2, P3, P4, P5, P6, P7, P8, PO3, PO4, PO7, PO8, O1, O2, FPz, AFz, Fz, FCz, Cz, CPz, Pz, POz, and Oz) and the right mastoid. The left mastoid was used as the reference site, and a ground electrode was placed on participants' right cheek. All electrodes were rereferenced offline to the average of both mastoids. Two electrodes placed above and below the left eye, as well as two electrodes placed on the outer canthi of both eyes, were used to record vertical and horizontal EOGs. EEG and EOG signals were amplified using a BrainAmp DC amplifier (Brain Products) and recorded at a sampling rate of 500 Hz.

MATLAB (The MathWorks, Inc., Natick, MA) and EEGLAB 13.02.2b free toolbox (Delorme & Makeig, 2004) were used for offline analysis.

EEG data were rereferenced offline to the average of both mastoids and filtered with a 1- to 30-Hz passband. Epochs of 600 msec before and after the onset of the visual outcome were extracted from the continuous EEG. Baseline correction was performed using the average voltage in a 100-msec time window preceding outcome onset. Epochs contaminated with large artifacts were identified and excluded whenever the voltage on an individual channel exceeded 400 μV to remove epochs with extremely large voltage fluctuations and whenever the joint probability of a trial exceeded 5 *SD*s to remove epochs with improbable data (Delorme, Sejnowski, & Makeig, 2007). To correct the remaining artifacts, the data were subjected to a temporal independent component analysis (Makeig, Bell, Jung, & Sejnowski, 1996; Jutten & Herault, 1991) using the infomax algorithm (Bell & Sejnowski, 1995). The resulting component matrix was screened for independent components (ICs) representing stereotyped artifact activity, such as horizontal (saccade) and vertical (blink) eye movements. This was done using a multistep correlational template-matching process as implemented in CORRMAP v1.02 (Viola et al., 2009). Topographies of ICs labeled as artifacts by the CORRMAP procedure were visually inspected and then calculated out of the data using inverse matrix multiplication. To increase signal-to-noise ratio for the analyses of the mediofrontal negativity (Debener, 2005), frontocentral ICs corresponding to the mediofrontal negativity were extracted from the data. To this end, we first chose an IC that showed a typical frontocentral distribution from one participant and then correlated all ICs of all participants with this template. The component with the highest correlation was individuated, and all components with a correlation within the range of .1 from the highest were selected and averaged. Note that cleaning the data from artifacts beforehand prevented that ICs related to artifacts could fall within this range. The mean number of components selected for each participant was 3.04 ($SD = 2.16$), and the mean correlation with the template was .65 ($SD = .05$). The presence of the typical frontocentral distribution was checked. These ICs were selected and back-projected into channel space using inverse matrix multiplication. All mediofrontal negativity analyses were performed using these back-projected data.

The mediofrontal negativity was quantified using a baseline independent peak-to-peak measure. As done in numerous previous studies (e.g., Frank, Woroch, & Curran, 2005; Yeung & Sanfey, 2004), we preferred this method over an area measure because it does not assume equivalent baselines immediately preceding the mediofrontal negativity and thus avoids distortions of the component (see San Martin, 2012, for an overview of Methods and Discussion). To obtain the peak-to-peak amplitude, the difference between the minimum peak in the 275- to 350-msec time window after outcome onset and the maximum peak in the 200- to 275-msec time window after outcome onset was computed (e.g., Ferdinand et al., 2012; Holroyd, Hajcak, & Larsen, 2006). To avoid a possible influence of differing trial numbers, the conditions were matched for trial number. To this end, for each participant, the condition with the smallest trial number was identified. This number of trials was then randomly drawn from each of the remaining conditions for 1000 times. Peak-to-peak amplitudes of the mediofrontal negativity were calculated for each iteration and condition and then averaged over all iterations separately for each condition. All statistical analyses were performed with RStudio v0.98.1062 (Boston, MA).

SCR Recording and Analysis

Ag–AgCl electrodes (model TSD203; Biopac Systems, Goleta, CA), filled with isotonic hyposaturated conductant, were used for recording SCRs. These were attached to participants' volar surface of the index and middle fingertips in their left hand (which did not require any motor movements during the task). A DC amplifier (Biopac GSR100) was used while recording the SCRs. A gain factor of 5 μS/V and low-pass filter set at 10 Hz were used for recording the analog signal, which was then passed through a MP-150 digital converter at a 200-Hz rate. The digital signal was then fed into AcqKnowledge 3.9 (Biopac Systems) and transformed into microsiemens for offline analysis. Data were analyzed offline using custom-made MATLAB scripts (The MathWorks, Inc.), and all statistical analyses were performed with RStudio v0.98.1062. All trials were recorded; however, for the analyses, only 88 catch trials in which no shock was delivered were considered to exclude shock artifacts.

SCR was extracted from the continuous signal and calculated for each trial as the peak-to-peak amplitude of the largest deflection during the 0.5- to 4.5-sec time window after stimulus onset (Schiller et al., 2008). The minimal response criterion was 0.02 μS, and smaller responses were encoded as zero. Raw SCR scores were square root-transformed to normalize the distributions and scaled to each participant's maximal US response to account for interindividual variability (Schiller et al., 2008). Three participants were excluded because of a malfunctioning of the SCR recorder instrument, and four participants had nonmeasurable levels of skin conductance. Thus, the SCR analysis included 41 participants.

RESULTS

SCR Results

SCR responses during shock (CS+) and neutral (CS−) trials were compared to test for implicit acquisition of the Pavlovian contingencies.

A 2 × 2 mixed-effect model was performed, with Task (early/late outcome) as the between-participant independent variable, CS (CS+/CS−) as the within-participant independent variable, and SCR as the dependent variable. A significant main effect of CS ($F(1, 39) = 4.11$, $p = .05$, part. $\eta^2 = .09$) was found, with CS+ trials (mean = 0.1 μS, SD = 0.06 μS) presenting higher SCR levels relative to CS− trials (mean = 0.089 μS, SD = 0.05 μS; Figure 2). All other effects were not significant (ps > .54).

Thus, participants' SCRs revealed implicit learning of the association between the CS (CS+ or CS−) and their respective outcome (shock or neutral), as higher levels of SCR were found during CS+ trials as compared with CS− trials.

Furthermore, explicit learning was assessed by asking the participant to pair each stimulus with its own outcome, at the end of each block. The CSs were correctly indicated 99.79% of the time.

EEG Results

The aim of the EEG analysis was to contrast and compare the amplitude of mediofrontal negativity elicited by shock and neutral outcomes delivered at expected and unexpected timings. Statistical analyses were performed on the data from the frontocentral electrode FCz, where mediofrontal negativity has been reported (Hajcak, Moser, Holroyd, & Simons, 2006). A mixed-effect model was performed, using Task (early/late outcome) as the between-participant independent variable and Timing (expected/unexpected) and Salience (shock/neutral) as the within-participant independent variables. Peak-to-peak amplitude was used as the dependent variable (see EEG Recording and Analysis). Results showed a significant main effect of Salience ($F(1, 46) = 18.88$, $p < .001$, part. $\eta^2 = .29$), qualified by a significant Task × Salience interaction ($F(1, 46) = 9.3$, $p < .001$, part. $\eta^2 = .17$), and a significant main effect of Timing ($F(1, 46) = 15.84$, $p < .001$, part. $\eta^2 = .26$), qualified by a significant Salience × Timing interaction ($F(1, 46) = 7.17$, $p = .01$, part. $\eta^2 = .13$; Figure 3). All other effects were not statistically significant (ps > .17).

Bonferroni-corrected post hoc analysis on the Task × Salience interaction revealed an overall significant difference ($p < .01$) between shock trials (mean = 3.62 μV, SD = 2.79 μV) and neutral trials (mean = 2.91, SD = 2.44) in the late-outcome task, but not in the early-outcome task ($p = 1$; shock trials: mean = 2.46 μV, SD = 1.73 μV; neutral trials: mean = 2.34 μV, SD = 1.61 μV).

Bonferroni-corrected post hoc analysis on the Salience × Timing interaction revealed a significant difference (ps < .01) between shock at expected timing (mean = 2.77 μV, SD = 2.07 μV) and shock at unexpected timing (mean = 3.30 μV, SD = 2.65 μV) and between neutral (mean = 2.68, SD = 2.17 μV) and shock at unexpected timing. No difference was found between neutral at expected timing (mean = 2.57 μV, SD = 2.02 μV) and neutral at unexpected timing ($p = 1$) and between neutral and shock at expected timing ($p = .5$).

Critically, this analysis evidenced the presence of significantly higher peak-to-peak mediofrontal negativity amplitude for salient (aversive) outcomes delivered earlier or later than expected (shock unexpected), as compared with both salient outcomes delivered at the expected timing (shock expected) and neutral outcomes delivered earlier or later than expected (neutral unexpected; Figure 3). Thus, salient outcomes unexpectedly shifted in time selectively trigger stronger mediofrontal negativity.

Grand-averaged scalp topographies showed a frontocentral distribution of peak-to-peak mediofrontal negativity amplitudes for all timing and salience conditions, with a maximum activation for unexpected timing of shock outcomes (Figure 3).

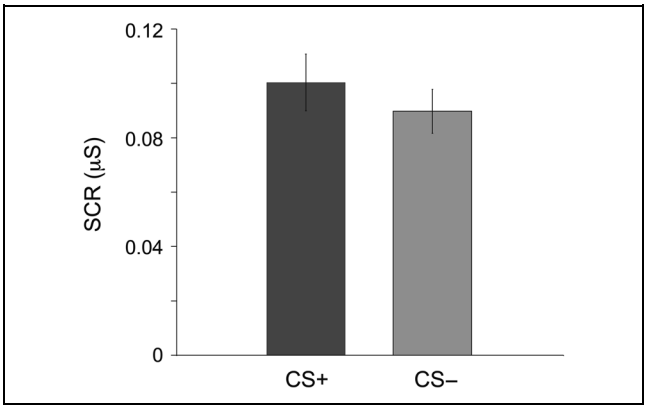

Figure 2. SCR levels at stimulus presentation during the Pavlovian aversive conditioning task. Bars represent *SEM*.

DISCUSSION

Learning about the time of events in the environment is a fundamental adaptive behavior (Cohen, 2011; Buhusi & Meck, 2005). Like other types of associative learning, it is deemed to depend critically on detecting mismatches (i.e., prediction errors) between expected and actual experiences (Holroyd & Coles, 2002; Sutton & Barto, 1998; Pearce & Hall, 1980; Mackintosh, 1975; Rescorla & Wagner, 1972). By recording ERPs while manipulating salience and participants' timing expectation of outcomes during classical aversive conditioning, this study aimed at investigating whether the mediofrontal negativity is sensitive to discrepancies in the expected and actual timing of events. The principal findings of the research may be summarized as follows: First, the amplitude of mediofrontal negativity was significantly higher for outcomes

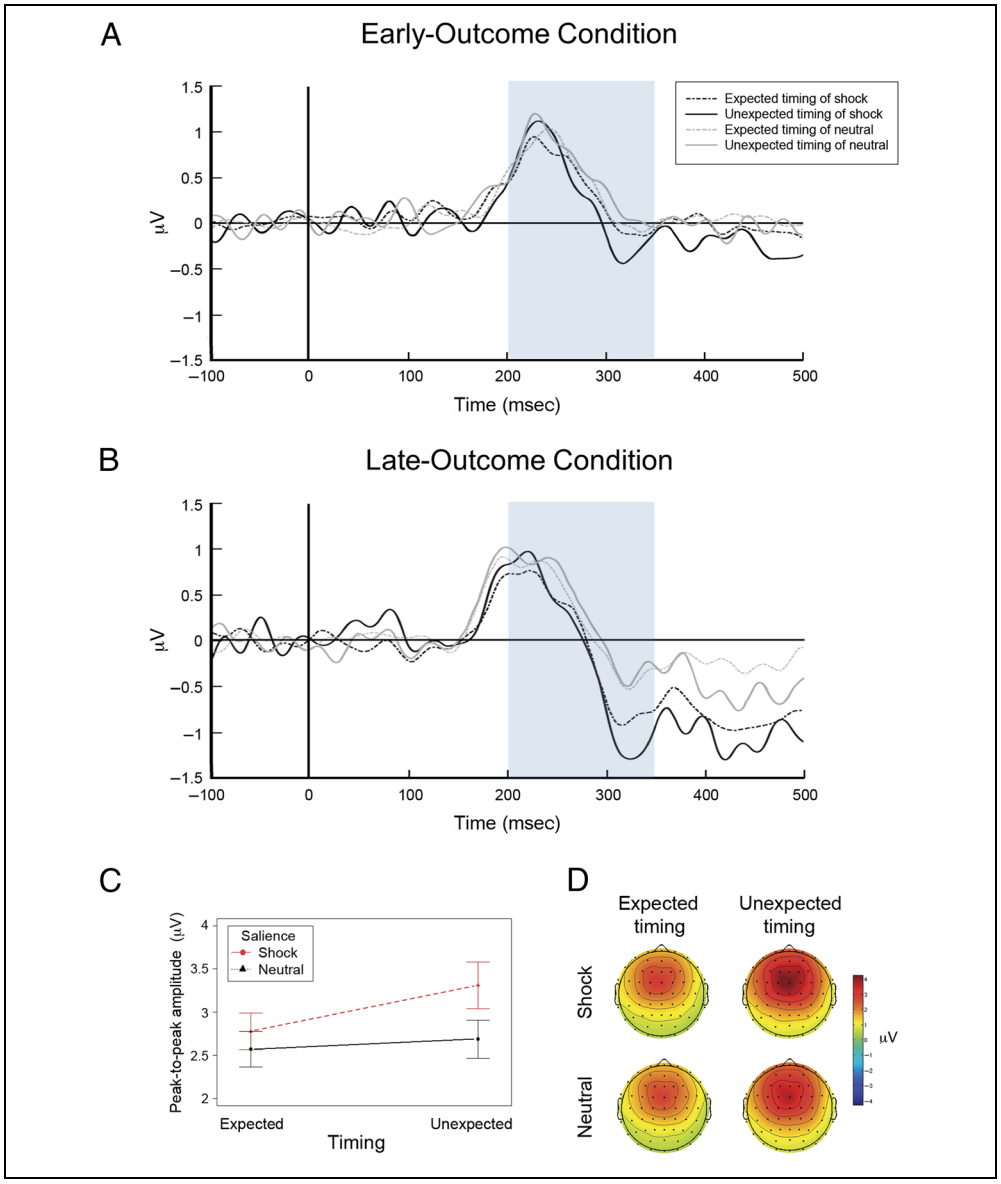

Figure 3. ERPs. A and B show the grand-averaged ERP waveforms from channel FCz in the early- and late-outcome conditions, respectively. The colored bands indicate the considered time window (200–350 msec); 0 represents the visual outcome onset. C shows the average peak-to-peak amplitude in all timing (expected/unexpected) and salience (shock/neutral) conditions. D shows the scalp distribution of the peak-to-peak amplitude of the 200- to 350-msec time window after visual outcome onset.

delivered at unexpected, relative to expected, times, even if the timing of the outcome was completely irrelevant to participants' behavior during the task. Second, the peak of the mediofrontal negativity was observed over central and frontocentral sites. Importantly, scalp negativities similar in timing, appearance, and scalp topography were observed when the shocks were delivered (1.5 sec) earlier or later than usual. Third, the effect of timing was obtained only for salient, not for neutral, outcomes. In this respect, the analysis of both explicit and implicit measures of conditioning confirmed that participants correctly associated predictive stimuli (CS+/CS−) to their respective outcome (shock/neutral), thereby indicating that they acquired an accurate representation of outcome salience.

Overall, these findings provide novel evidence about neural activity in ACC/mPFC as it is reflected in the mediofrontal negativity. They reveal that this component tracks the timing of salient events and reports an error signal when an aversive outcome delivery is shifted away from an expected time.

Although some experimental work has attempted to address whether the mediofrontal negativity is modulated by the timing of outcome, no prior studies, to our knowledge, have examined whether this ERP component is sensitive to the temporal aspect of outcome prediction by showing prediction error responses with surprising changes (i.e., unexpectedness) in outcome timing, as proposed by the PRO model.

In previous studies, temporal expectations or predictions were not varied: Either external cues signaled in advance whether timing preceding outcome delivery would be short or long (Peterburs, Kobza, & Bellebaum, 2016; Wang, Chen, Lei, & Li, 2014) or short and long outcome timings occurred in 50% of cases and thus were equally expected (Weinberg, Luhmann, Bress, & Hajcak, 2012). By contrast, the current study directly manipulated

the probability of occurrence of an outcome at a given time (i.e., temporal expectation), thus making it possible to contrast expected and unexpected timing information.

The present findings are in line with those of fMRI studies reporting activity in frontal regions and VTA modulated by predictions of the timing of outcome delivery (Forster & Brown, 2011; Klein-Flügge et al., 2011). Crucially, the observed pattern of mediofrontal negativity for outcomes occurring at unexpected times is consistent with the predictions of the PRO model that ACC/mPFC signals violations of predicted outcome timing (Alexander & Brown, 2011, 2014). Like many previous models (Matsumoto, Matsumoto, Abe, & Tanaka, 2007; Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006; Rudebeck, Walton, Smyth, Bannerman, & Rushworth, 2006), the PRO model posits that mPFC can learn to predict outcomes by using a reinforcement learning algorithm based on prediction errors (surprising outcomes). However, in contrast to previous accounts, the PRO model also maintains that the mPFC encodes multiple independent predictions in parallel, which concerns not only the likelihood of occurrence of outcomes but also their timing. Consistent with the PRO model predictions, we found that the scalp potentials overlying ACC/mPFC activity were significantly more negative when outcomes occurred at unexpected times, even if the outcomes themselves were expected, suggesting that ACC/mPFC reports errors in the prediction of the timing of outcomes independently of their probability of occurrence.

Given that mPFC is thought to play a critical role in action selection (Holroyd & Coles, 2002; Posner & Petersen, 1990), earlier formulation of the PRO model focused on the role of mPFC in predicting the outcome of an action (Alexander & Brown, 2011). However, to account for several findings implicating mPFC in paradigms not explicitly related to response generation, simulation works (Alexander & Brown, 2014, 2015) have proposed an extended PRO model of mPFC, according to which this region may encode predictions and prediction errors even when the predicted outcomes are not contingent on prior actions. In close agreement with the extended implementation of the PRO model, the present findings reveal that the mediofrontal negativity encodes the unexpected timing of outcomes during a task with no action requirement (i.e., Pavlovian aversive conditioning). These results appear consistent with a variety of empirical evidences suggesting that ACC/mPFC is critically implicated in learning the causal structure of the environment (Dayan & Niv, 2008), allowing an individual to predict what follows what and when.

Notably, the PRO model maintains that mPFC learns to predict and signal discrepancies between expected and actual events of any stimulus and outcome a participant may encounter during the course of an experimental task. In striking contrast with this prediction, however, the current results reveal that ACC/mPFC activation is sensitive to the salience of the outcome, as a higher amplitude of mediofrontal negativity was selectively observed for aversive outcomes occurring at unexpected timings, as compared with neutral outcomes similarly delivered at unexpected timings. These findings are highly consistent with earlier EEG works (Lou et al., 2015; Talmi et al., 2013; Maier et al., 2011, 2012; Ganushchak & Schiller, 2008; Hajcak et al., 2005), suggesting that the mediofrontal negativity expresses a prediction error related to the motivational salience of outcomes.

It is worth noting that the way salience is defined may vary between studies. Whereas, in some studies, the term "salience" defines the magnitude of the outcome and provides a measure of the degree of motivation or drive with which the outcome is approached or avoided (Kahnt, Park, Haynes, & Tobler, 2014; Talmi et al., 2013; Matsumoto & Hikosaka, 2009; Roesch & Olson, 2004), in others, salience refers to a large category of arousing and alerting events that signal the need for cognitive and behavioral changes (Zink, Pagnoni, Martin-Skurski, Chappelow, & Berns, 2004; Horvitz, 2000; Redgrave, Prescott, & Gurney, 1999), including and extending beyond reinforcing events. Unfortunately, this study fails to separate magnitude and arousal-based concepts of salience, because aversive stimuli are both arousing and reinforcing events. Regardless of the particular definition, however, the salience of an event determines the amount of attentional resources and control processes that are engaged to mobilize the most adaptive responses (i.e., behavioral, autonomic, and cognitive; see Ham, Leff, de Boissezon, Joffe, & Sharp, 2013; Shackman et al., 2011; Seeley et al., 2007). In this context, aversive events (as electrical shocks used in the present experiment) can be regarded as salient, because the defensive preparatory adjustments elicited can help an organism to minimize the impact of negative consequences (Ploghaus, Becerra, Borras, & Borsook, 2003). Therefore, learning the precise timing of delivery of a salient outcome constitutes a critical feature of an accurate prediction, whereas the timing of occurrence of a neutral outcome does not have the same relevance. Interestingly, the modulation of mediofrontal negativity by motivational salience was more pronounced in the late-outcome task than in the early-outcome task (regardless of unexpectedness), presumably because of greater temporal contiguity between the CS presentation and US delivery in the late-outcome task (average CS–US interval = 1300 msec) than in the early-outcome (average CS–US interval = 2200 msec) task (Rescorla & Cunningham, 1979).

A noteworthy aspect of this study is the finding of the same pattern of mediofrontal negativity both when the aversive shock was delivered earlier and later than its usual times. In the early-outcome condition, the expectation of outcome is relatively low at the beginning of the interval and reaches its peak at 2.5 sec (when the outcome is most likely to occur). Consequently, all events occurring before that time are unexpected and should produce a temporal prediction error, which nicely accords with the present data. On the other hand, in

the late-outcome condition, the expectation peaks 1 sec after the predicting (CS) stimulus and, if the outcome fails to occur, grows monotonically at each successive time step during the elapsing interval (i.e., "the objective hazard rate"—the probability that an outcome will occur at a given time given that it has not yet occurred). Therefore, later-than-average outcome events are expected and should produce no temporal prediction error (Daw, Courville, & Tourtezky, 2006; Coull & Nobre, 1998), which contradicts the present results. To account for this seeming discrepancy, it must be noted, however, that our experimental design included a proportion of trials in which the outcome failed to occur (i.e., catch trials). The presence of catch trials implies that later-than-average outcomes cannot be confidently expected when they do not occur at the habitual time of 1 sec. When the usual time of 1 sec had passed, participants may have relaxed their outcome expectation in anticipation of a catch trial. Surprise at the appearance of a late shock would then augment sharply the mediofrontal negativity, as we reported here. In line with this interpretation, an fMRI study (Klein-Flügge et al., 2011) revealed VTA activation in response to unexpectedly late rewards only when participants were uncertain over whether an outcome would be delivered on a trial. Thus, we conjecture that uncertainty with respect to the occurrence of outcome plays a critical role in modulating the mediofrontal negativity elicited by a delayed outcome. Additional work, however, will be needed to test this proposal directly.

To conclude, many current theories view mPFC as crucially implicated in processing outcomes and signaling discrepancies between actual and expected outcomes (Alexander & Brown, 2011; Holroyd & Coles, 2002). The present findings on the mediofrontal negativity, an ERP component generated by activity in ACC and adjacent mPFC, provide further evidence for this view. We found that the mediofrontal negativity is sensitive to the timing of salient outcomes by reporting prediction error responses with surprising changes in outcome timing. These signals may be used to learn about temporal contingencies in the environment and make preparatory adjustments in advance of the occurrence of motivationally significant events (Shackman et al., 2011). These findings represent a critical step toward reaching a clearer understanding of ACC/mPFC activity and the cognitive processes indexed by the mediofrontal negativity.

Acknowledgments

The authors thank Giulia Petrillo for helping with data acquisition. This work was supported by grants from the University of Bologna (Ricerca Fondamentale Orientata) and from the Ministero Istruzione Università e Ricerca (PRIN 2010, protocol no. 2010XPMFW4_009) to G. d. P.

Reprint requests should be sent to Sara Garofalo, Department of Psychiatry, University of Cambridge, Forvie Site, Addenbrooks, Cambridge CB2 2QQ, United Kingdom, or via e-mail: sg732@cam.ac.uk.

REFERENCES

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience, 14,* 1338–1344.

Alexander, W. H., & Brown, J. W. (2014). A general role for medial prefrontal cortex in event prediction. *Frontiers in Computational Neuroscience, 8,* 1–11.

Alexander, W. H., & Brown, J. W. (2015). Hierarchical error representation: A computational model of anterior cingulate and dorsolateral prefrontal cortex. *Neural Computation, 27,* 2354–2410.

Amador, N., Schlag-Rey, M., & Schlag, J. (2000). Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *Journal of Neurophysiology, 84,* 2166–2170.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7,* 1129–1159.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience, 6,* 755–765.

Cohen, M. X. (2011). It's about time. *Frontiers in Human Neuroscience, 5,* 1–15.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience, 18,* 7426–7435.

Daw, N. D., Courville, A. C., & Tourtezky, D. S. (2006). Representation and timing in theories of the dopamine system. *Neural Computation, 18,* 1637–1677.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology, 18,* 185–196.

Debener, S. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25,* 11730–11737.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage, 34,* 1443–1449.

Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *Journal of Neuroscience, 32,* 12087–12092.

Forster, S. E., & Brown, J. W. (2011). Medial prefrontal cortex predicts and evaluates the timing of action outcomes. *Neuroimage, 55,* 253–265.

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts report reinforcement learning and conflict biases. *Neuron, 47,* 495–501.

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *Neuroimage, 39,* 395–405.

Garofalo, S., Maier, M. E., & di Pellegrino, G. (2014). Mediofrontal negativity signals unexpected omission of aversive events. *Scientific Reports, 4,* 1–7.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology, 71,* 148–154.

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology, 42,* 151–160.

Ham, T., Leff, A., de Boissezon, X., Joffe, A., & Sharp, D. J. (2013). Cognitive control and the salience network: An investigation of error processing and effective connectivity. *Journal of Neuroscience, 33,* 7091–7098.
Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., et al. (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. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.
Holroyd, C. B., Hajcak, G., & Larsen, J. T. (2006). The good, the bad and the neutral: Electrophysiological responses to feedback stimuli. *Brain Research, 1105,* 93–101.
Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience, 96,* 651–656.
Jessup, R. K., Busemeyer, J. R., & Brown, J. W. (2010). Error effects in anterior cingulate cortex reverse when error likelihood is high. *Journal of Neuroscience, 30,* 3467–3472.
Jutten, C., & Herault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing, 24,* 1–10.
Kahnt, T., Park, S. Q., Haynes, J.-D., & Tobler, P. N. (2014). Disentangling neural representations of value and salience in the human brain. *Proceedings of the National Academy of Sciences, U.S.A., 111,* 5000–5005.
Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J., & Rushworth, M. F. S. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience, 9,* 940–947.
Klein-Flügge, M. C., Hunt, L. T., Bach, D. R., Dolan, R. J., & Behrens, T. E. J. (2011). Dissociable reward and timing signals in human midbrain and ventral striatum. *Neuron, 72,* 654–664.
Lou, B., Hsu, W. Y., & Sajda, P. (2015). Perceptual salience and reward both influence feedback-related neural activity arising from choice. *Journal of Neuroscience, 35,* 13064–13075.
Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.
Maier, M. E., di Pellegrino, G., & Steinhauser, M. (2012). Enhanced error-related negativity on flanker errors: Error expectancy or error significance? *Psychophysiology, 49,* 899–908.
Maier, M. E., & Steinhauser, M. (2013). Updating expected action outcome in the medial frontal cortex involves an evaluation of error type. *Journal of Neuroscience, 33,* 15705–15709.
Maier, M. E., Yeung, N., & Steinhauser, M. (2011). Error-related brain activity and adjustments of selective attention following errors. *Neuroimage, 56,* 2339–2347.
Makeig, S. J., Bell, A., Jung, T.-P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems, 8,* 145–151.
Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature, 459,* 837–841.
Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience, 10,* 647–656.
Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience & Biobehavioral Reviews, 28,* 441–448.
Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research, 171,* 213–224.
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,* 1994–2004.
Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.
Peterburs, J., Kobza, S., & Bellebaum, C. (2016). Feedback delay gradually affects amplitude and valence specificity of the feedback-related negativity (FRN). *Psychophysiology, 53,* 209–215.
Ploghaus, A., Becerra, L., Borras, C., & Borsook, D. (2003). Neural circuitry underlying pain modulation: Expectation, hypnosis, placebo. *Trends in Cognitive Sciences, 7,* 197–200.
Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.
Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short latency dopamine burst too short to signal reinforcement error? *Trends in Neurosciences, 22,* 146–151.
Rescorla, R. A., & Cunningham, C. L. (1979). Spatial contiguity facilitates Pavlovian second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 5,* 152.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
Roesch, M. R., & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science, 304,* 307–310.
Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. S. (2006). Separate neural pathways process different decision costs. *Nature Neuroscience, 9,* 1161–1168.
Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience, 11,* 389–397.
Sambrook, T. D., & Goslin, J. (2016). Principal components analysis of reward prediction errors in a reinforcement learning task. *Neuroimage, 124,* 276–286.
San Martin, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience, 6,* 304.
Schiller, D., Cain, C. K., Curley, N. G., Schwartz, J. S., Stern, S. A., Ledoux, J. E., et al. (2008). Evidence for recovery of fear following immediate extinction in rats and humans. *Learning and Memory, 15,* 394–402.
Schultz, W., & Dickinson, A. (2000). Neural coding of prediction errors. *Annual Review of Neuroscience, 23,* 473–500.
Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience, 27,* 2349–2356.
Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience, 12,* 154–167.
Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: Single neuronal signals related to degree of reward expectancy. *Science, 296,* 1709–1711.

Silvetti, M., Seurinck, R., & Verguts, T. (2013). Value and prediction error estimation account for volatility effects in ACC: A model-based fMRI study. *Cortex, 49,* 1627–1635.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience, 33,* 8264–8269.
Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology, 120,* 868–877.
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,* 1870–1884.
Wang, J., Chen, J., Lei, Y., & Li, P. (2014). P300, not feedback error-related negativity, manifests the waiting cost of receiving reward information. *NeuroReport, 25,* 1044–1048.
Weinberg, A., Luhmann, C. C., Bress, J. N., & Hajcak, G. (2012). Better late than never? The effect of feedback delay on ERP indices of reward processing. *Cognitive, Affective & Behavioral Neuroscience, 12,* 671–677.
Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience, 24,* 6258–6264.
Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron, 42,* 509–517.