Frontal Theta Links Prediction Errors to Behavioral Adaptation in Reinforcement Learning

James F. Cavanagh[1,†], **Michael J. Frank**[2], **Theresa J. Klein**[1], and **John J.B. Allen**[1]
[1]University of Arizona, Tucson, AZ

[2]Brown University, Providence, RI

Abstract

Investigations into action monitoring have consistently detailed a fronto-central voltage deflection in the Event-Related Potential (ERP) following the presentation of negatively valenced feedback, sometimes termed the Feedback Related Negativity (FRN). The FRN has been proposed to reflect a neural response to prediction errors during reinforcement learning, yet the single trial relationship between neural activity and the quanta of expectation violation remains untested. Although ERP methods are not well suited to single trial analyses, the FRN has been associated with theta band oscillatory perturbations in the medial prefrontal cortex. Medio-frontal theta oscillations have been previously associated with expectation violation and behavioral adaptation and are well suited to single trial analysis. Here, we recorded EEG activity during a probabilistic reinforcement learning task and fit the performance data to an abstract computational model (Q-learning) for calculation of single-trial reward prediction errors. Single-trial theta oscillatory activities following feedback were investigated within the context of expectation (prediction error) and adaptation (subsequent reaction time change). Results indicate that interactive medial and lateral frontal theta activities reflect the degree of negative and positive reward prediction error in the service of behavioral adaptation. These different brain areas use prediction error calculations for different behavioral adaptations: with medial frontal theta reflecting the utilization of prediction errors for reaction time slowing (specifically following errors), but lateral frontal theta reflecting prediction errors leading to working memory-related reaction time speeding for the correct choice.

Keywords

FRN; Reinforcement Learning; Prediction Error; Theta; Anterior Cingulate

Investigations into action monitoring have consistently detailed a fronto-central voltage deflection in the Event-Related Potential (ERP) following the presentation of negatively valenced feedback, sometimes termed the Feedback Related Negativity (FRN). A leading theory of the FRN suggests that it is reflective of the degree of negative reward prediction error (Holroyd and Coles, 2002) – that is, the degree to which outcomes are worse than expected. However, alternative evidence suggests that the variance in feedback-locked ERPs are primarily due to positive prediction errors (Holroyd et al., 2008). It is possible that

†Corresponding Author Department of Psychology, University of Arizona 1503 University Ave, Tucson, AZ (USA) *Tel:* (520) 621-7086 jim.f.cav@gmail.com.

shortcomings inherent to the ERP methodology, including cross-trial averages and difference waves, have contributed to an opaque account of feedback related neuroelectric activities. Compounding this dilemma, difficulty in quantifying reward expectation may have led to untested assumptions. Here, we quantified reward expectation using computational models of reinforcement learning. These computational values were used to interrogate medio-frontal theta band oscillatory perturbations (arguably, the basis of the FRN) at a single trial level. In this report, we present evidence that interactive medial and lateral frontal theta activities reflect the degree of reward prediction error in the service of behavioral adaptation. Moreover, both positive and negative prediction error are reflected in frontal theta, but different brain areas use these calculations for different behavioral adaptations.

Similarities between the eliciting circumstances of the FRN and the functioning of the mesolimbic dopamine system (Schultz, 2002) have yielded an influential theoretical account of FRN generation based on reinforcement learning principles (Holroyd and Coles, 2002). Reinforcement learning theory suggests that the processes underlying the ability to learn to seek reward and avoid punishment in an uncertain environment can occur through trial and error, by using the difference between expected outcomes and external feedback to incrementally update internal representations of state-action values (Sutton and Barto, 1998). The Holroyd and Coles (2002) reinforcement learning theory of the FRN postulates that the response-locked Error Related Negativity (ERN) and the stimulus locked FRN (which they term fERN) are reflections of the same generic high level error processing system, and that activation of feedback- and response-related systems are inversely related as learning progresses from reliance on external stimuli (larger FRN) to reliance on internal representations (larger ERN). This reinforcement learning account specifically suggests that the FRN is reflective of the computation of negative reward prediction error – a signature of when events are worse than expected (Holroyd and Coles, 2002; Holroyd et al., 2004; Holroyd et al., 2003; Nieuwenhuis et al., 2004a; Nieuwenhuis et al., 2004b).

One direct prediction of the reinforcement learning theory of the FRN is that *single-trial* variations in amplitude should reflect the degree of negative prediction error (Holroyd and Coles, 2002; Nieuwenhuis et al., 2004a), a postulate that has not been directly tested yet, possibly due to the cross-trial averaging procedure common to ERPs. Feedback from any condition that is not optimal, such as not gaining the highest amount when expecting inevitable gain, elicits an FRN (Holroyd et al., 2004; Nieuwenhuis et al., 2004b). The FRN is larger to unexpected or infrequent negative feedback (Cohen et al., 2007; Donkers et al., 2005; Holroyd et al., 2003; Potts et al., 2006; Yasuda et al., 2004; but see: Cohen et al., 2007), fitting with a reinforcement learning account. However, FRN amplitude is not sensitively modulated by the magnitude of negative outcome between conditions (Gehring and Willoughby, 2002; Hajcak et al., 2006; Holroyd et al., 2004; Marco-Pallares et al., 2008; Yeung and Sanfey, 2004). Parametric changes in expectation of loss (three or more conditions) have been reflected by incrementally larger FRN amplitudes (Holroyd and Coles, 2002; Nieuwenhuis et al., 2002; Holroyd et al., 2009), although this effect is sometimes minor (Holroyd et al., 2004) or non-existent (Hajcak et al., 2005), unless participants are primed to define their expectation (Hajcak et al., 2007). These discrepancies in parametric estimation and the absence of magnitude dependent modulation suggest that it is necessary to estimate the participant's expectation in order to accurately predict FRN amplitude dependencies. Computational models of reinforcement learning that fit individual participant's trial-by-trial sequence of choices can provide reasonable estimates of these expectations.

Another important determinant of FRN magnitude is whether behavioral adaptation is possible, and if so, whether negative feedback can be used to alter behavior (Cohen and Ranganath, 2007; Hajcak et al., 2005; Holroyd et al., 2009; Holroyd et al., 2003; Yasuda et al., 2004; Yeung et al., 2005). This sensitivity to decision and action suggests that the FRN is intimately

related to the utilization of negative information in the service of behavioral adaptation. Indeed, Cohen and Ranganath (2007) have shown that within-subjects, larger FRN amplitudes precede behavioral switches and this pattern qualitatively fits a computational simulation that used prediction errors to guide future behavioral choice. Furthermore, across-subjects, individual differences in FRN magnitude are predictive of the degree to which participants subsequently avoid decisions with negative outcomes (Frank et al., 2005). Variations in the morphology and amplitude of the FRN across studies indicate that the FRN is maximally sensitive to feedback eliciting a negative prediction error in the service of future behavioral adaptation, despite its reliable occurrence when outcomes are worse than expected more generally.

Although the prevailing literature focuses on the sensitivity of the FRN to negative feedback, a recent study suggests that the major differences in ERPs during reinforcement learning occur on correct trials (Holroyd et al., 2008). Motivated by previous fMRI and EEG studies (Nieuwenhuis et al., 2005; van Veen et al., 2004), these authors argued that the FRN reflects the same underlying processes as that ERP component associated with perceptual mismatch in an oddball paradigm, the N2, with which the FRN shares many similarities in terms of eliciting conditions, scalp topography, and timing (Donkers et al., 2005; Holroyd, 2002). This account suggests that a voltage positivity exists on better-than-expected trials and that occurs in lieu of the FRN/N2. Indeed, a voltage positivity following correct feedback has been empirically observed and is sensitive to reinforcement-learning contingencies of events being better than expected (Eppinger et al., 2008; Holroyd et al., 2008; Potts et al., 2006). It is clear that a formal investigation of prediction error in relation to both positive and negative feedback is necessary to begin to sort out these differing, and sometimes conflicting accounts of the EEG responses to reinforcement cues.

One under-addressed issue in the FRN literature is the limitation imposed by the ERP signal averaging methods. A growing literature suggests that ERP components such as the FRN may be reflective of stimulus-driven phase realignment and power increases of ongoing oscillatory activity, rather than a singular 'burst' event (Fell et al., 2004; Le Van Quyen and Bragin, 2007; Makeig et al., 2004; Makeig et al., 2002; Sauseng et al., 2007). While ERPs may not always be generated by the alteration of ongoing oscillations, the methodological means to parse these generative circumstances are fraught with ambiguity (Ritter and Becker, 2009; Sauseng et al., 2007; Yeung et al., 2004; Yeung et al., 2007). Although one need not adopt an oscillatory view to examine activity at the single trial level, this perspective has the potential to provide novel insights into neurocognitive function as well as allowing methodological advancements that are not assessable by the ERP method, such as characterization of single trial activities and changes in presumed functional communication between brain areas.

Both the ERN and the FRN have been shown to reflect a degree of theta phase consistency and power enhancement over the medial frontal cortex (Bernat, et al., 2008; Cavanagh et al., 2009; Cohen et al., 2007; Marco-Pallares et al., 2008; Trujillo and Allen, 2007), supporting the major postulate of Holroyd and Coles' (2002) reinforcement learning theory that these two ERPs reflect the same generic high level error processing system. We recently provided evidence that the medial PFC (mPFC) error processing system interacts with lateral PFC (lPFC) cognitive control systems following response errors via theta band phase synchrony (Cavanagh et al., 2009). A separate study also found theta band phase synchrony between mPFC and lPFC, which increased linearly with increasing conflict during a Stroop task (Hanslmayr et al., 2008). These sort of network-wide coherent oscillations are thought to reflect entrained inter-regional activity, increasing the coordination of spike timing across spatially separate neural networks and presumably reflecting functional communication (Buzsáki, 2006; Buzsaki and Draguhn, 2004; Fries, 2005; Womelsdorf et al., 2007). Theta oscillations may represent a general operating mechanism of medial and lateral frontal cortices involved in action monitoring and behavioral adjustment.

In sum, the FRN has been proposed to reflect the degree of negative prediction error, but crucial aspects of this theory remain untested: particularly the quantification of expectation to allow trial-by-trial correlations between FRN and prediction error. Another account suggests that the FRN does not reflect the degree of negative prediction error; rather, positive prediction errors are reflected by other ERP components that act to obscure the N2/FRN. Both of these accounts may be hindered by reliance on the ERP method of averaging over total ongoing voltage activities. We propose that the FRN is at least partially reflective of theta band oscillatory perturbations in the mPFC that are intimately related to expectation violation, behavioral adaptation, and interaction with lPFC cognitive control systems. To test these differing and sometimes conflicting accounts, we investigated EEG activity during a probabilistic reinforcement learning task. EEG data were first converted to current source density to diminish volume condition, and then decomposed using time/frequency methods (wavelet convolution and the Hilbert transform) for investigation of single trial theta band power and phase relations. Performance data from the reinforcement learning task were fit to an abstract computational model (Q-learning; (Sutton and Barto, 1998), which estimated action values and prediction errors, providing a quantification of the degree to which events are better or worse than expected. We present evidence that interactive medial and frontal theta activities reflect the degree of prediction error in the service of behavioral adaptation following both positive and negative feedback.

Methods

Participants

All participants gave informed consent and the research ethics committee of the University of Arizona approved the study. Participants were recruited based upon pretest materials given to undergraduate students in introductory psychology classes. Participants were invited to a screening session if they indicated low levels of depressive symptomatology on the Beck Depression Inventory (BDI score <7) during the pretest. The screening session was used to identify participants who fit the recruitment criterion for the EEG session: 1) aged 18 – 25, 2) stable low BDI (<7) and no self-reported history of major depressive disorder, 3) no current psychoactive medication use, 4) no history of head trauma or seizures, 5) no self-reported symptoms indicating a possibility of an Axis I disorder, as indicated by self-reported computerized completion of the Electronic Mini International Neuropsychiatric Interview (eMINI: Medical Outcome Systems, Jacksonville, FL). All participants received experimental credit for their participation in screening and EEG sessions. A total of 75 participants were recruited for the EEG session reported in this study, although additional exclusion criterion reduced the final number of participants included in this study to 50 (see below).

Probabilistic Learning Task

The probabilistic learning task consisted of a forced choice training phase consisting of up to six blocks of sixty trials each, followed by a subsequent testing phase (Frank et al., 2004). During the training phase the participants were presented with three stimulus pairs, where each stimulus was associated with a different probabilistic chance of receiving 'Correct' or 'Incorrect' feedback. These stimulus pairs (and their probabilities of reward) were termed A / B (80% / 20%), C / D (70% / 30%) and E / F (60% / 40%). Over the course of the training phase, a participant usually learns to choose A over B, C over D and E over F, solely due to adaptive responding based on the feedback.

All training trials began with a jittered inter-trial-interval between 300 and 700 ms. The stimuli then appeared for a maximum of 4000 ms, and disappeared immediately after the choice was made. If the participant failed to make a choice within the 4000 ms, "No Response Detected" was presented. Following a button press, either 'Correct' or 'Incorrect' feedback was presented

for 500 ms (jittered between 50 and 100 ms post response). Although this timing may allow an overlap in the EEG between later parts of the response with early stimulus processing, immediate feedback may be necessary for adequate encoding of the action value in the basal ganglia (Frank et al., 2005; Maddox et al., 2003).

The participants underwent training trials (consisting of one to six blocks of sixty stimuli each) until they reached a minimum criterion of choosing the probabilistically best stimulus in each pair (AB ≥ 65%, CD ≥ 60%, and EF ≥ 50% correct choices). This same criterion has been used in multiple prior studies with this task. Participants who did not reach this criterion by the end of the sixth block were moved to the testing phase regardless. During the testing phase all possible stimulus pairs (ie. AD, CF, etc.) were presented eight times (120 trials total) and no feedback was provided. Data from the test phase were only analyzed if participants selected the most rewarding stimulus (A) over the least rewarding (most negative) stimulus (B) more than 50% of the time when this stimulus pair was presented during the testing phase, since data from participants who fail this basic criterion are not interpretable (Frank et al., 2007a; Frank et al., 2004; Frank et al., 2005). This criterion removed 14 participants from the analyses. In addition to the exclusion criterion for behavioral performance, participants were excluded if there were fewer than 30 EEG epochs in any condition (this excluded an additional 11 participants). Data are only presented from the training phase of the experiment in this report, as feedback was only presented during the training phase. Note that the use of the terms "correct" and "incorrect" throughout refer to the feedback, not to the optimal or accurate response (i.e. incorrect refers to the "Incorrect" feedback given on an 'A' (80% correct) stimulus, even though this was the high probability or optimal choice).

Abstract Computational Modeling of Performance Data

The trial-by-trial sequence of choices for each subject was fit by a Q-learning reinforcement learning model (Sutton and Barto, 1998; Watkins, 1992). Q-learning assigns expected reward values to actions taken during a particular state (i.e. choosing A when seeing an A / B pair). We refer to these state-action values as Q values. See Figure 1 for a depiction of Q learning. As in Frank et al. (2007b), this model includes separate learning rate parameters for gain and loss (correct and incorrect) feedback trials in the training phase of the probabilistic learning task. These separate gain/loss learning rates (αG/αL) scaled the updating of the stimulus-action values separately for rewards and punishments. The expected value (Q) of any stimulus (i) at time (t) was computed after each reinforcer (R=1 for Correct, R=0 for Incorrect):

$$Q_i(t+1) = Q_i(t) + \alpha G[R(t) - Q_i(t)]_+ + \alpha L[R(t) - Q_i(t)]_-,$$

where αG and αL are learning rates from gains and losses, respectively, and which are multiplied by prediction errors to update Q values. These Q values were entered into a softmax logistic function to produce probabilities (P) of responses for each trial, with higher probabilities predicted for stimuli having relatively larger Q values, and a free parameter for inverse gain (β) to reflect the tendency to explore or exploit:

$$P_i(t) = \exp\left(Q_i(t)/\beta\right) / \Sigma j \exp\left(Q_j(t)/\beta\right),$$

These probabilities are then used to compute the log likelihood estimate (LLE) of the subject having chosen that set of responses for a given set of parameters over the whole training phase. The parameters that produce the maximum LLE were found using the Simplex method, a standard hill-climbing search algorithm implemented with Matlab (The MathWorks, Natick, MA) function 'fmincon'.

$$LLE = \sum_t \ln\left(P_i\left(t\right)\right),$$

Since the best-fit learning rates and exploration parameters will differ between subjects, they will result in somewhat different predictions errors and performance. By choosing parameters that maximize the likelihood of producing the actions selected by the subject, the resultant model fits each individual subject. The parameters that produced the maximum log likelihood are selected, and the Q-values produced by this particular set of parameters were saved. Prediction errors (PE) for each subject were then computed on a trial-by-trial basis from the estimated Q value of the chosen stimulus at that time:

$$PE = R\left(t\right) - Q_i\left(t\right),$$

These prediction errors were then examined in a trial-by-trial fashion with the stimulus-locked EEG data. Characterization of model fits were computed as Pseudo-R^2 statistics ((LLE − r) / r, where r is the LLE of a chance performance model (Camerer and Ho, 1999; Frank et al., 2007b)). Two participants had poor fits to the Q learning model (no change in initial parameters, R^2=0) and parameters were unable to be derived for prediction error. Grand-averaged model parameters and fits are reported without these two participants. See Frank et al. (2007b) for further details on Q-learning model fits.

EEG Recording and Pre-Processing

Scalp voltage was measured using 62 Ag/AgCl electrodes referenced to a site immediately posterior to Cz. Additionally, two mastoid channels were recorded, as were separate bipolar channels for recording horizontal and vertical eye movements. EEG was recorded continuously in AC mode with bandpass filter (0.5-100 Hz) with a sampling rate of 500Hz. Impedances were kept under 10 kΩ. All raw EEG data were visually inspected by two researchers to reject bad sections with artifacts and to identify bad channels to be interpolated. Data were then epoched (−1500ms to 1500ms) around each feedback onset in the training phase, these epochs were then cleaned of eyeblink and muscle artifacts using Independent Components Analysis from the EEGLab toolbox (Delorme and Makeig, 2004). Epochs were then transformed to Current Source Density (CSD) using the methods and functions of Kayser and Tenke (2006) (and also algebraically re-referenced to linked mastoids for analyses in the supplemental materials). CSD computes the second spatial derivative of voltage between nearby electrode sites, acting as a reference-free spatial filter. The CSD transformation highlights local electrical activities at the expense of diminishing the representation of distal activities (volume conduction). The diminishment of volume conduction effects by CSD transformation may reveal subtle local dynamics and also lead to more accurate characterization of local activities during the calculation of long-distance synchrony. See Figure 1 for CSD ERPs **(filtered from .5 to 15 Hz)**.

Supplementary figures (S1-S3) contrast linked mastoid voltage ERPs, CSD ERPs, theta band filtered CSD ERPs, and Hilbertized CSD theta power for correct, incorrect, and difference wave conditions. These plots highlight potential differences in condition-wide expectancy (which could contribute to component overlap) as in Holroyd and Krigolson (2007). There was no evidence for component overlap due to change in condition-wide expectancy or due to volume conduction in theta band filtered CSD-EEG, suggesting that these spatial and temporal filters effectively isolate phase-locked mPFC theta proposed to underlie the FRN. To assess whether this activity was specific to medial-frontal theta, and not in part influenced by later theta from parietal P3-related processes, we examined the distribution of CSD power over

the interval from 200-400 ms. There was a remarkably consistent mid-frontal scalp topography, with no evidence of posterior contributions that would be expected if P3-related processes where contributing to the signal. (see supplementary figure S4).

Grand Average Time-Frequency Calculations

Time-frequency calculations were computed using custom-written Matlab routines (Cavanagh et al., 2009; Cohen et al., 2008) of the method described by Lachaux et al. (1999). The CSD-EEG time series in each epoch was convolved with a set of complex Morlet wavelets, defined as a Gaussian-windowed complex sine wave: $e^{-i2\pi t f} e^{-t^2/(2*\sigma^2)}$, where t is time, f is frequency (which increased from 1 to 50Hz in 50 logarithmically spaced steps), and σ defines the width (or "cycles") of each frequency band, set according to $4.5/(2\pi f)$. This convolution resulted in: 1) estimates of instantaneous power (the magnitude of the analytic signal), defined as Z[t] (power time series: $p(t) = real[z(t)]^2 + imag[z(t)]^2$); and, 2) phase (the phase angle) defined as $\phi_t = arctan(imag[z(t)]/real[z(t)])$. Each epoch was then cut in length (−500 to +1000 ms peri-feedback) and baseline corrected to the average frequency power in each condition from −300 to −200 ms prior to the onset of the feedback. Power was normalized by conversion to a decibel (dB) scale (10*log10[power(t)/power(baseline)]), allowing a direct comparison of effects across frequency bands.

Two different types of oscillation phase coherence were examined: Inter-Trial Phase Coherence (ITPC) and Inter-Channel Phase Synchrony (ICPS). For convenience, we use the term 'coherence' when describing the consistency of phase angles over trials within a single electrode (ITPC), and the term 'synchrony' when describing the consistency of phase angles between two channels (ICPS), even though we do not assess the existence of zero phase difference for a textbook definition of 'synchrony'. ITPC measures the consistency of phase values for a given frequency band at each point in time over trials, in one particular electrode. Phase coherence values vary from 0 to 1, where 0 indicates random phases at that time/frequency point across trials, and 1 indicates identical phase values at that time/frequency point across trials. The phase coherence value is defined as:

$$ITPC = |\frac{1}{n} * \sum_{x=1}^{n} e^{i\phi_{xt}}|,$$

where n is the number of trials for each time and each frequency band. ITPC thus reflects the extent to which oscillation phase values are consistent over trials at that point in time-frequency space (power, in contrast, represents the intensity of that signal).

ICPS measures the extent to which oscillation phases are similar across different electrodes over time/frequency. ICPS is calculated in a similar fashion as inter-trial phase coherence:

$$ICPS = |\frac{1}{n} * \sum_{t=1}^{n} e^{i[\phi_{jt} - \phi_{kt}]}|,$$

where n is the number of trials, ϕ_j and ϕ_k are the phase angles of electrode j and k. Thus, phase angles are extracted from two electrodes, and then subtracted: If the phase angles from the two electrodes fluctuate in synchrony over a period of time, their difference will be constant (i.e., nonuniformly distributed), leading to ICPS values close to 1.

ITPC and ICPS values were computed as the percent change from the pre-cue baseline. Note that wavelet convolution necessarily "smears" activity over time at the expense of better frequency resolution. Values for statistical analysis were averaged over time and frequency in

windows defined by the grand average wavelet plots in Figure 2 (over the theta band (4-8 Hz), averaged across 200-500 ms, separately correct and incorrect trials). Topographical plots of the incorrect minus correct difference display these averaged values for each separate metric. Lateral electrodes of interest (F5/6) were selected based on a previous investigation (Cavanagh et al., 2009); subsequent analyses did not reveal any qualitative or quantitative difference between other lateral electrode pairs, as described below.

Single Trial Time-Frequency Calculations

Single trial analyses of power and phase synchrony were computed using the Hilbert transform on EEG data filtered in the theta band from 5-8 Hz (using a zero phase-shift filter), separately for the FCz (mPFC), F5 (left lPFC) and F6 (right lPFC) electrodes. A smaller theta band width was determined based on temporal constraints, allowing a minimum of 2.5 cycles of the lowest frequency (5 Hz). Individual matrices were sorted according to the degree of prediction error (Y axis) separately for correct and incorrect feedback trials, which were then cut in length (X axis: −500 to +1000 ms peri-feedback), and interpolated into a common time-frequency space for grand averaging; this method shows the amount of theta power (Z axis; "cold" to "hot" colors) for increasing degrees of prediction error (these figures are sometimes referred to as ERPimages), see Figure 3. Measurement windows for single trial values were derived based on windows defined by the wavelet-based analyses, the ERPimages and mathematical constraints for the number of cycles required for phase synchrony estimates. Theta band power values were averaged in window of 200 to 400 ms post feedback for each trial. Theta phase synchrony was calculated in a window of 100 to 600 ms post feedback for each trial (at least 2.5 cycles of theta).

Post-feedback Reaction Time (RT) changes were computed as the trial-to-trial difference in milliseconds (ms) between trials (subsequent trial RT minus the current trial RT), with higher values reflecting greater slowing. This RT difference was computed for two different types of conditions: 1) "**delay**", the next trial type of the same stimulus pair (i.e. the next AB pair following an AB pair, with a varied number of interspersing CD or EF trials), and 2) "immediate", the subsequent trial of an opposite stimulus type (i.e. the next stimulus following an AB pair, as long as it was a CD or EF pair). For statistical analysis, mixed linear models (MLMs) were used (via SAS Proc Mixed). MLMs estimate the relationship between two variables (as beta weights) within each participant in the first level; here we investigated single-trial level variables including PFC activities (phase synchrony or power), prediction error (coded as an absolute value), or reaction time change (immediate or delayed). The second level may include moderating or condition-wide variables; here we included accuracy (correct, incorrect) and in the case of lPFC analyses, hemisphere (left, right). Note that prediction error was coded as an absolute value in MLMs with an additional level for accuracy; this approach allowed assessment of both the unsigned magnitude of prediction error (main effect) and the direction of prediction error (interaction with accuracy). Outliers exceeding three standard deviations around each individual variable mean were removed prior to analyses **(~1% of trials were removed on average)**. Mixed model results are reported as unstandardized β, standard error, t statistic (β/SE), and p value. Figure 4 shows the individual standardized beta weights and standard errors for the relationships of interest. For display purposes, select predictor variables were median split and the values of the dependent variable were averaged within the upper and lower bins (see Figure 5).

Results

Demographics and Performance

Participants (N=50, 26 female) were an average of 19 years old (SD= 1.35). All participants were right handed (as assessed by the scale of Chapman & Chapman, 1987). Participants

completed an average of four blocks of training (SD= 1.6), and were 65% accurate (SD= 9%) in the Test phase. There were an average of 127 EEG epochs of correct feedback (SD= 48.6) and 103 epochs of incorrect feedback (SD= 45.7) per participant. The average RT on trials with incorrect feedback was 1145ms (SD= 364). RT on trials with correct feedback was 1132ms (SD= 353); there was no difference in RT between these conditions (t<1). The average learning rate for correct trials (αG) was .41 (SD= .35), for incorrect trials (αL) was .19 (SD= .25), and the average inverse gain (β) parameter was .38 (SD= .31). Average Log Likelihood for the training fit was −144 (SD= 73), average Pseudo R^2 was .18 (SD= .15).

Immediate and Delayed Behavioral Adaptation Following Feedback

Following positive reinforcement ('correct'), participants chose the same stimulus again the next time it appeared 74% (SD= 11%) of the time (a "win-stay" strategy). Following punishment ('incorrect'), participants were nearly evenly split in their choice during the next stimulus pairs, switching only 45% (SD= 8%) of the time ("lose-switch"). In contrast to the design of Cohen and Ranganath (2007), this more complex reinforcement learning task required the slow integration of negative feedback (where punishment cues are only partially reflective of the true stimulus value); a strong lose-switch strategy throughout the entire task would ultimately be counterproductive. This pattern was evident in participant behavior, where switches following negative feedback occurred less than half of the time.

There were no aggregate effects of behavioral adaptation in immediate RT change initially following negative feedback (post-incorrect M= 0.2ms, SD= 89; p>.9), nor was there a notable RT change initially following positive feedback (post-correct M= 16ms, SD= 67; p=.10). By contrast, for the next repetition of a given stimulus after a variable number of intervening items (delay RT), there was a general effect of post-error slowing and post-correct speeding for **delay** RT change (post-incorrect M= 26.89ms, SD= 52.89; post-correct M= −23.81, SD= 37.50; each was significantly different than 0, and significantly different from each other, p's< .01), indicating an aggregate effect of post-error slowing and post-correct speeding for repeated stimulus pairs of the same type.

Delay RT changes occurred both in conjunction with the utilization of feedback and in the decision to switch or stay. **Delay** RT was slower for switches after both types of feedback (correct-switch M= 25ms, SD= 169; incorrect-switch M= 15ms, SD= 12), but there was dissociation between accuracy conditions where speeding occurred on correct-stay trials (M= -35ms, SD= 54) yet slowing occurred on incorrect-stay trials (M= 32ms, SD= 90). Although the 2-way interaction for these accuracy-behavior combinations did not reach the standard threshold for statistical significance ($F_{(1,49)}$= 3.59, p= .06), the general trend of RT slowing for **delay RT** effects with the exception of RT speeding following correct-stay trials helps to interpret dissociation between accuracy conditions in single-trial brain behavior relationships.

Grand Average Theta to Feedback

As expected, incorrect feedback trials had significantly greater mPFC theta power (FCz: $t_{(49)}$= 6.47, p<.01) and mPFC theta phase coherence (FCz: $t_{(49)}$= 2.8, p<.01) than did correct trials (see Figure 2)[1]. A 2 (Accuracy: Correct, Incorrect) * 2 (Hemisphere: F5, F6) ANOVA for mPFC-lPFC theta synchrony revealed that theta synchrony was greater following incorrect compared to correct feedback (Accuracy: $F_{(1,49)}$= 7.74, p<.01) especially over right hemisphere (Interaction: $F_{(1,49)}$= 8.84, p<.01; absent a main effect for Hemisphere p> .47), see Figure 2.

---

[1]However, neither mPFC theta power nor lPFC-mPFC theta synchrony predicted the degree that a participant learned to avoid decisions associated with negative outcomes or to choose decisions associated with positive outcomes as assessed during the test phase (a NoGo or Go learning bias; p's> .24), see (Frank et al., 2004). Thus, these frontal theta activities appear to reflect the rapid trial-to-trial adaptation following feedback presumed to be computed by PFC, not the slow, incremental, integrative, and potentially implicit, reinforcement learning processes of the basal ganglia (Frank et al., 2007b).

A similar 2 * 2 ANOVA for power between left and right lPFC revealed main effects for hemisphere and accuracy without an interaction effect (Accuracy: $F_{(1,49)}$= 6.0, p<.05; Hemisphere: $F_{(1,49)}$= 5.0, p<.05, Interaction: $F_{(1,49)}$= 3.23, p=.08), where power was greater after incorrect feedback in the presence of an overall right hemispheric bias. There were no significant accuracy or hemispheric effects for theta band phase coherence in lPFC sites (p> .16). Averaged theta activities did not predict averaged future behavior when switching or staying with the same stimulus choice the next time it was shown following incorrect or correct feedback (neither power at FCz, F6 or F5, nor synchrony between F6-FCz and F5-FCz, all p's > .11). In sum, mPFC (power and phase coherence) and right lPFC (power and phase synchrony with mPFC) demonstrated greater theta band activities following incorrect feedback, but neither of these effects related to averaged trends in behavioral adaptation.

Single Trial Theta, Behavior, and Model-Derived Prediction Errors

As detailed in the methods section, mixed linear models (MLMs) were used to assess the single-trial bivariate relationships between pairings of PFC activities, prediction error, and reaction time change as a function of accuracy (and in the case of lPFC analyses, hemisphere). Quantitative depictions of theta power, prediction error and reaction time change relationships are shown in Figure 4 as standardized beta weights; qualitative depictions and raw data values are shown in Figure 5.

Prediction Error Predicts Immediate RT Change

In a model predicting immediate RT change [RT = PE*Accuracy], the magnitude of prediction error was a significant predictor (β= 164, SE= 56, t= 2.9, p<.01), whereas accuracy was not (p>.13), with a significant interactive effect (β= −355, SE= 68, t= −5.23, p<.01) whereby reaction time was slower following increasingly negative prediction errors and was faster following increasingly positive prediction errors (Figure 4A). Note these RT effects were found in relation to PE despite the fact that only choices, and not RTs, were fit by the computational model. **Delay** RT change was not related to prediction error (all p's >.21).

mPFC Theta Scales with Negative Prediction Error

In a model predicting mPFC theta power [mPFC = PE*Accuracy], theta power over the mPFC was greater with higher absolute magnitude prediction error (β= 1.38, SE= .49, t= 2.84, p<. 01), and greater on incorrect trials (β= —2.28, SE= .11, t=−20.44, p<.01), with an interaction indicating that incorrect trials had the largest relationship between prediction error and mPFC theta power (β= −1.46, SE= .61, t= −2.38, p<.05). Negative prediction error directly scaled with mPFC theta power (β= 1.48, SE= .50, t= 2.96, p<.01), but positive prediction error did not (p>.8) (Figure 4B).

mPFC Theta Predicts Immediate RT Change

In a model predicting immediate RT change [RT = mPFC*Accuracy], RT slowing was predicted by greater medial PFC theta power (β= 4.17, SE= 1.64, t= 2.55, p<.01), with no accuracy or interaction effects (p's >.16) (Figure 4B). Immediate RT change was not predicted by mPFC-lPFC synchrony (p's > .11). **Delay** RT change was not related to mPFC theta power or mPFC-lPFC synchrony (all p's >.21).

Negative Prediction Error Does Not Predict Immediate RT Change in the Absence of mPFC Theta

To test whether the relationship between negative prediction error and immediate RT slowing [RT = −PE] is related to mPFC theta power, we accounted for the variance in immediate RT slowing due to mPFC theta prior to inclusion in an MLM with negative prediction error [RT_residual = −PE]. Negative prediction error directly predicted immediate RT slowing (β=

147, SE= 50, t= 2.92, p<.01) but negative prediction error no longer significantly predicted immediate RT slowing after accounting for variance in mPFC theta (β= .22, SE= .13, t= 1.7, p>.05) (Figure 5A).

lPFC Theta Scales with Absolute Prediction Error

In a model predicting lPFC theta power [lPFC = PE*Accuracy*Hemisphere], theta power over the lPFC was greater in the right hemisphere following feedback, as revealed by a PE*hemisphere interaction (β= .97, SE= .42, t= 2.32, p<.05), absent a main effect for accuracy (p >.14) (Figure 4D). Theta synchrony between mPFC and lPFC was also related to the magnitude of prediction error (β= .038, SE= .02, t= 1.99, p<.05) and was larger following incorrect feedback (β= −.01, SE= .005, t= −1.91, p=.05), but there was no PE*accuracy interaction (p>.13) or hemispheric specificity (p>.55) (Figure 4C).

lPFC Theta Predicts Delay RT Change

In a model predicting delay RT [RT = lPFC*Accuracy*Hemisphere], lPFC power predicted **delay** RT following correct trials specifically, where the RT for the same stimulus pair was faster (lPFC*Accuracy interaction: β= −6.24, SE= 3.03, t= −2.06, p<.05) without a main effect of lPFC power (p>.34) and without hemispheric specificity (p>.67) (Figure 4D). There were no relationships between lPFC power and immediate RT slowing (p's >.22).

These single-trial effects were robust across frontal regions. Alternative electrodes were investigated based on the scalp topographies in Figure 2: more ventral (F7/8) and more dorsal (AF3/4). The pattern of theta band relationships with prediction error and reaction time at these electrode sites was similar to F5/6. Although theta power on correct trials appears to peak earlier than incorrect trials (see Figure 3), no new findings were revealed using a 100-300 ms time window for single-trial analysis. Whereas previous investigations have noted beta band (17-25 Hz or 20-30 Hz) relationships to increasingly positive feedback (Cohen et al., 2007; Marco-Pallares et al., 2008), we did not find any relationships between beta activity (15-30 Hz) and prediction error or reaction time adaptation.

Discussion

This investigation revealed that theta band activities following reinforcement cues are related to both the magnitude of prediction violation and the degree of future trial-to-trial behavioral adaptation. This finding suggests that theta band activities in the frontal cortex are involved in the evaluation of positive and negative feedback in the service of learning and/or strategic adjustment. Moreover, these theta band activities are reflective of presumed functions of the underlying, presumably generative, cortices: evaluation of punishment and immediate behavioral adaptation in the medial PFC, and working memory in the dorsal areas of the lateral PFC. These findings may help clarify hypotheses advanced in the ERP literature that suggest a number of different, and sometimes contradictory accounts of the voltage deflection to negative feedback, the FRN.

Event-related theta band power and oscillatory perturbation occur with the same topography and time range as the FRN, and exhibit similar modulation to punishment cues (Bernat, et al., 2008; Cohen et al., 2007; Marco-Pallares et al., 2008). This perspective fits with other accounts suggesting that frontal midline theta activities underlie the ERN and predict subsequent behavioral adaptation following errors (Cavanagh et al., 2009; Debener et al., 2005; Luu et al., 2003; Luu et al., 2004; Trujillo and Allen, 2007). The oscillatory perspective would further imply that these distinct neuroelectric phenomena of the ERN and the FRN may reflect a similar underlying neural mechanism. Theta band activities may reflect general mechanism of

computation in the mPFC, which is modulated in response to punishment, error, and immediate behavioral adaptation.

Worse Than Expected

Theta activity in different cortical areas may reflect the utilization of prediction error in the service of behavioral adaptation. The magnitude of prediction error was linearly related to the degree of immediate reaction time adjustments: slower following increasingly negative prediction errors and faster following increasingly positive prediction errors. Medial PFC theta power covaried with both the degree of prediction error and immediate reaction time slowing following errors, fitting with the proposed function of the mPFC and Anterior Cingulate Cortex (ACC) in reacting to punishment and immediate behavioral adaptation (Blair et al., 2006; Bush et al., 2002; Shima and Tanji, 1998; Wrase et al., 2007). Synchrony between mPFC and lPFC sites was also related to the degree of prediction error and was greater following incorrect feedback, although this measurement did not specifically relate to reaction time adjustments. Medial PFC theta power may reflect the processes underlying the translation of prediction error for immediate behavioral slowing, especially following errors.

Varying accounts suggest that the FRN is reflective of the degree of negative reward prediction error (Holroyd and Coles, 2002), and of the utilization of information for future behavioral adaptation (Cohen and Ranganath, 2007; Holroyd et al., 2009). This current investigation supports these basic propositions about the FRN as indexed by theta band power. Bartholow et al. (2005) have additionally shown that between-condition alteration of expectancy can alter the amplitude of both the N2 and the ERN, suggesting that varied stimulus and response-related mPFC activities are also modulated by expectation. Although two previous studies have failed to show modulation of theta band power to error feedback in conditions that might result in increasingly negative prediction errors (Cohen et al., 2007; Marco-Pallares et al., 2008) these investigations may have been similarly hindered as previous studies of the FRN due to the lack of a quantitative estimation of prediction expectation and the lack of behavioral dependency (see: Holroyd et al. (2009)).

EEG source estimation of the FRN has implicated both the anterior and posterior cingulate cortices, as well as pre-supplemental motor areas (Cohen and Ranganath, 2007; Luu et al., 2003; Miltner et al., 1997; van Schie et al., 2004). The ACC has been proposed to encode reward prediction errors after feedback in order to update predictions for use in guiding future behaviors (Cohen, 2007; Holroyd and Coles, 2002). ACC activity has been shown to be reflective of prediction error in fMRI (Cohen, 2007), primate electrophysiology (Matsumoto et al., 2007; Shima and Tanji, 1998), and intracranial human EEG (Oya et al., 2005). The ACC and surrounding mPFC may play a role in determining the value of exercised options based on environmental feedback (Glascher et al., 2009), which may be used to drive changes in behavior when external cues indicate that new strategies are required (Paus, 2001). In experimental settings, the lose-switch strategy has been highlighted as a specific reaction in ACC to positive punishment in both fMRI (Wrase et al., 2007) and single cell recordings (Shima and Tanji, 1998).

The sensitivity of the ACC to punishers may be an inherent feature of the process of integrating internal and external cues in the service of future behavioral adaptation. These functions are thought to occur in parallel with the more incremental, integrative, and potentially implicit, reinforcement learning processes of the basal ganglia (Frank et al., 2007b). Punishers may indicate that behavior needs to be adapted in the future, whereas rewards may indicate that behavior is adequate - at least until the need for exploration of other rewards arises. A parsimonious account of the mechanisms that generate the FRN, and theta power as described here, may summarize these operations as being common to the processing demands of the

ACC, which is particularly sensitive to errors, behavioral selection, and the conjunction of the two.

Wang et al. (2005) have shown that theta activity is generated in Area 24 of the dorsal ACC, where extremely local areas respond to error, conflict, novelty, familiarity, and stimulus-response associations in the time range of the FRN and ERN. ACC generated theta activity has been shown in monkeys, being responsive to movement, reward, and lack of reward (Tsujimoto et al., 2006). Event-related theta power change and oscillatory perturbation may reflect the operations of the ACC and surrounding mPFC during outcome evaluation and action selection, especially when changing behavior following errors.

Better Than Expected

Participants tended to slow down following incorrect feedback and speed up following correct feedback when they were presented with the next stimulus pair of the same type. This **delay RT** effect was not seen on the stimulus pair immediately following feedback but on the next appearance of the same stimulus following some delay, so we interpret these effects as being partially reflective of working memory for the outcome specific to the stimulus chosen (Frank et al., 2007b). Reaction time speeding was specifically associated with correct-stay choices, and may reflect the tendency to encode the positive stimulus-action-outcome association in working memory so as to repeat the selection in the future. Although prediction error did not relate to the degree of **delay** RT change, both of these processes were associated with right lPFC power increases. Right lPFC theta power covaried with the degree of prediction error after both correct and incorrect feedback, indicating that this area might be involved in calculating the degree to which events are different than expected regardless of the valence of prediction error. Lateral PFC activities would thus be ideally suited for holding the value of the prediction over the delay for future behavioral adaptation, fitting with the presumed role of the lPFC in working memory (Braver et al., 1997).

This investigation did not find a unique signal over the mPFC when events were better than expected. Increasingly positive prediction errors and increased magnitude of reward have been associated with greater power and phase coherence in the beta band in the mPFC (Cohen et al., 2007; Marco-Pallares et al., 2008). We did not find any effects in the beta band in relation to prediction error or behavioral adjustment in this study. Holroyd et al. (2008) suggest that a positivity on correct trials occurs instead of an FRN/N2, and that this positivity is reflective of reinforcement contingencies. While we didn't observe any time-frequency activities at the FCz electrode that were greater on correct than incorrect trials, we did not investigate ERP effects that could be due to **slow, non-oscillatory potentials or to** component overlap. In fact, our approach sought to diminish any possible component overlap: the CSD transform and theta band filter act as spatial and temporal filters, as demonstrated in the supplemental materials. Previous application of the CSD transform has indicated that response-locked positivities preceding errors (Ridderinkhof et al., 2003) may actually be reflective of a stimulus-locked positivity that is volume conducted over diminished response-locked negativities (Cavanagh et al., 2009; Vidal et al., 2003). While a similar volume conduction effect could be occurring during correct-related positivities, the present investigation is not suited to probing this effect and we remain agnostic about the relevance this ERP component. Future investigations may wish to directly identify a better-than-expected voltage component and then seek to determine the oscillatory characteristics and topography of the underlying event-related EEG.

Mismatch and Prediction Error

Technically, a system that responds to the magnitude of negative and positive prediction error in the same direction (such as lPFC theta) doesn't compute a "reward prediction error", which by definition requires a valenced signal (e.g., see axiomatic description by Caplin and Dean

(2008). The absolute value of a prediction error, which seems to be reflected in lPFC theta, may reflect something akin to salience of expectation violation. ACC activity has been shown to be reflective of both positive and negative prediction error (Matsumoto et al., 2007), even though sensitivity to punishment and uncertainty may particularly characterize ACC function (Blair et al., 2006; Bush et al., 2002; Shima and Tanji, 1998; Wrase et al., 2007). It remains to be seen if mPFC theta (or the FRN) is the actual reflection of **mesolimbic dopamine-driven** negative reward prediction computation, or if this signal is simply reflective of a general tendency of late ERPs and the underlying oscillatory perturbations to reflect mismatch (Friston, 2005), which is generated by cortex that is particularly sensitive to punishment and error. This perspective suggests that the cortical activities underlying the FRN/mPFC theta could occur separate or prior to midbrain dopamine nuclei activities (Frank et al., 2005), as opposed to being generated by them (Holroyd and Coles, 2002).

Parallels between the FRN and the mismatch N2 have been frequently noted in the literature, particularly due to the spatio-temporal pattern of these ERPs and the similar modulating factors. As with the FRN, infrequency and degree of mismatch also modulate the N2, however, it is unknown if these eliciting events reflect alterations of different underlying neural processes (Donkers et al., 2005; Donkers and van Boxtel, 2004; Holroyd, 2002). Holroyd et al. (2008) suggest that the FRN is simply an N2 that occurs to unexpected negative feedback, and that difference wave ERP variance depends more on positive than negative prediction errors. We suggest that both the FRN and the N2 are partial reflections of mPFC theta oscillatory perturbations. Here, we demonstrate that mPFC theta power following incorrect feedback is a reflection of a system that uses negative reward prediction errors to adapt future behavior (in line with the predictions of Holroyd and Coles (2002)).

All of the fronto-central negativities peaking ~250 ms post-stimulus appear to be sensitive to a form of expectation mismatch, although this type of prediction error may differ in terms of action selection (control N2), attention (mismatch N2) and negative reward prediction (FRN) (see Folstein and Van Petten (2008) for a review of control and mismatch N2). While these mismatch signals may reflect disparate processes in unique cognitive circumstances, all of these aforementioned processes have been specifically associated with ACC function, and these mismatch signals may be similarly reflected by mPFC theta oscillations. A parsimonious account may surmise that mechanisms generating scalp-recorded fronto-central voltage negativities are common to the processing demands of the mPFC, especially the ACC, which are varied across systems related to cognitive and motor control, attention, and reinforcement learning: but are especially active when using punishment and error signals to adapt future behavior.

Conclusion

This investigation supports the theory that the medio-frontal theta band activities, which presumably underlie the FRN component, are reflective of the degree of negative prediction error and subsequent behavioral adaptation. Moreover, multiple neural systems may be involved in the computation of different types of prediction error and the utilization of feedback for different behavioral adaptations. Theta band oscillations may be reflective of these prediction error calculations in the medial PFC for immediate behavioral adaptation and in lateral PFC for delayed behavioral adaptation.

Supplementary Material

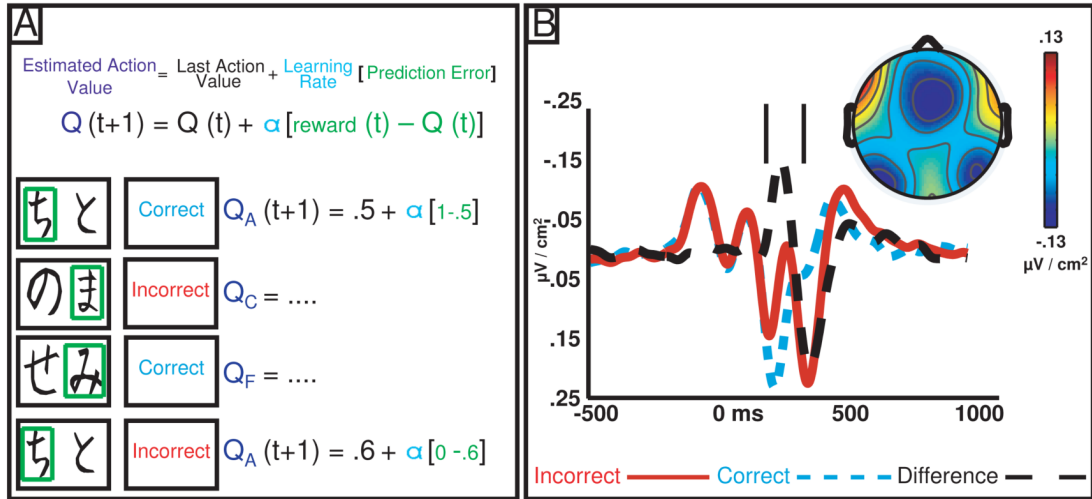Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bartholow BD, Pearson MA, Dickter CL, Sher KJ, Fabiani M, Gratton G. Strategic control and medial frontal negativity: beyond errors and response conflict. Psychophysiology 2005;42:33–42. [PubMed: 15720579]

Bernat EM, Nelson LD, Holroyd CB, Gehring WJ, Patrick CJ. Separating cognitive processes with principal components analysis of EEG time-frequency distributions. Proc. SPIE 2008;7074:70740S.

Blair K, Marsh AA, Morton J, Vythilingam M, Jones M, Mondillo K, Pine DC, Drevets WC, Blair JR. Choosing the lesser of two evils, the better of two goods: specifying the roles of ventromedial prefrontal cortex and dorsal anterior cingulate in object choice. Journal of Neuroscience 2006;26:11379–11386. [PubMed: 17079666]

Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC. A parametric study of prefrontal cortex involvement in human working memory. Neuroimage 1997;5:49–62. [PubMed: 9038284]

Bush G, Vogt BA, Holmes J, Dale AM, Greve D, Jenike MA, Rosen BR. Dorsal anterior cingulate cortex: A role in reward-based decision making. Proceedings of the National Academy of Sciences of the United States of America 2002;99:523–528. [PubMed: 11756669]

Buzsáki, G. Rhythms of the brain. Oxford University Press; Oxford ; New York: 2006.

Buzsaki G, Draguhn A. Neuronal oscillations in cortical networks. Science 2004;304:1926–1929. [PubMed: 15218136]

Camerer C, Ho TH. Experience-weighted attraction learning in normal form games. Econometrica 1999;67:827–874.

Caplin A, Dean M. Axiomatic methods, dopamine and reward prediction error. Current Opinion in Neurobiology 2008;18:197–202. [PubMed: 18678251]

Cavanagh JF, Cohen MX, Allen JJ. Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. J Neurosci 2009;29:98–105. [PubMed: 19129388]

Cohen MX. Individual differences and the neural representations of reward expectation and reward prediction error. Soc Cogn Affect Neurosci 2007;2:20–30. [PubMed: 17710118]

Cohen MX, Elger CE, Ranganath C. Reward expectation modulates feedback-related negativity and EEG spectra. Neuroimage 2007;35:968–978. [PubMed: 17257860]

Cohen MX, Ranganath C. Reinforcement learning signals predict future decisions. J Neurosci 2007;27:371–378. [PubMed: 17215398]

Cohen MX, Ridderinkhof KR, Haupt S, Elger CE, Fell J. Medial frontal cortex and response conflict: evidence from human intracranial EEG and medial frontal cortex lesion. Brain Res 2008;1238:127–142. [PubMed: 18760262]

Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 2005;25:11730–11737. [PubMed: 16354931]

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134:9–21. [PubMed: 15102499]

Donkers FC, Nieuwenhuis S, van Boxtel GJ. Mediofrontal negativities in the absence of responding. Brain Res Cogn Brain Res 2005;25:777–787. [PubMed: 16249075]

Donkers FC, van Boxtel GJ. The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. Brain Cogn 2004;56:165–176. [PubMed: 15518933]

Eppinger B, Kray J, Mock B, Mecklinger A. Better or worse than expected? Aging, learning, and the ERN. Neuropsychologia 2008;46:521–539. [PubMed: 17936313]

Fell J, Dietl T, Grunwald T, Kurthen M, Klaver P, Trautner P, Schaller C, Elger CE, Fernandez G. Neural bases of cognitive ERPs: more than phase reset. J Cogn Neurosci 2004;16:1595–1604. [PubMed: 15601521]

Folstein JR, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 2008;45:152–170. [PubMed: 17850238]

Frank MJ, D'Lauro C, Curran T. Cross-task individual differences in error processing: neural, electrophysiological, and genetic components. Cogn Affect Behav Neurosci 2007a;7:297–308. [PubMed: 18189003]

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 2007b;104:16311–16316. [PubMed: 17913879]

Frank MJ, Seeberger LC, O'Reilly R,C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 2004;306:1940–1943. [PubMed: 15528409]

Frank MJ, Woroch BS, Curran T. Error-related negativity predicts reinforcement learning and conflict biases. Neuron 2005;47:495–501. [PubMed: 16102533]

Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci 2005;9:474–480. [PubMed: 16150631]

Friston K. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 2005;360:815–836. [PubMed: 15937014]

Gehring WJ, Willoughby AR. The medial frontal cortex and the rapid processing of monetary gains and losses. Science 2002;295:2279–2282. [PubMed: 11910116]

Glascher J, Hampton AN, O'Doherty JP. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex 2009;19:483–495. [PubMed: 18550593]

Hajcak G, Holroyd CB, Moser JS, Simons RF. Brain potentials associated with expected and unexpected good and bad outcomes. Psychophysiology 2005;42:161–170. [PubMed: 15787853]

Hajcak G, Moser JS, Holroyd CB, Simons RF. The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. Biol Psychol 2006;71:148–154. [PubMed: 16005561]

Hajcak G, Moser JS, Holroyd CB, Simons RF. It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. Psychophysiology 2007;44:905–912. [PubMed: 17666029]

Hanslmayr S, Pastotter B, Bauml KH, Gruber S, Wimber M, Klimesch W. The electrophysiological dynamics of interference during the Stroop task. J Cogn Neurosci 2008;20:215–225. [PubMed: 18275330]

Holroyd, CB. A note on the Oddball N200 and the Feedback ERN. In: Ullsperger, MF,M., editor. Errors, Conflicts and the Brain. Current Opinions on Performance Monitoring. Dortmund, Germany: 2002. p. 211-218.

Holroyd CB, Coles MG. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 2002;109:679–709. [PubMed: 12374324]

Holroyd CB, Krigolson OE. Reward prediction error signals associated with a modified time estimation task. Psychophysiology 2007;44:913–917. [PubMed: 17640267]

Holroyd CB, Krigolson OE, Baker R, Lee S, Gibson J. When is an error not a prediction error? An electrophysiological investigation. Cogn Affect Behav Neurosci 2009;9:59–70. [PubMed: 19246327]

Holroyd CB, Larsen JT, Cohen JD. Context dependence of the event-related brain potential associated with reward and punishment. Psychophysiology 2004;41:245–253. [PubMed: 15032989]

Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD. Errors in reward prediction are reflected in the event-related brain potential. Neuroreport 2003;14:2481–2484. [PubMed: 14663214]

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;45:688–697. [PubMed: 18513364]

Kayser J, Tenke CE. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. Clin Neurophysiol 2006;117:348–368. [PubMed: 16356767]

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. Hum Brain Mapp 1999;8:194–208. [PubMed: 10619414]

Le Van Quyen M, Bragin A. Analysis of dynamic brain oscillations: methodological advances. Trends Neurosci 2007;30:365–373. [PubMed: 17559951]

Luu P, Tucker DM, Derryberry D, Reed M, Poulsen C. Electrophysiological responses to errors and feedback in the process of action regulation. Psychol Sci 2003;14:47–53. [PubMed: 12564753]

Luu P, Tucker DM, Makeig S. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. Clin Neurophysiol 2004;115:1821–1835. [PubMed: 15261861]

Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. J Exp Psychol Learn Mem Cogn 2003;29:650–662. [PubMed: 12924865]

Makeig S, Debener S, Onton J, Delorme A. Mining event-related brain dynamics. Trends Cogn Sci 2004;8:204–210. [PubMed: 15120678]

Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ. Dynamic brain sources of visual evoked responses. Science 2002;295:690–694. [PubMed: 11809976]

Marco-Pallares J, Cucurell D, Cunillera T, Garcia R, Andres-Pueyo A, Munte TF, Rodriguez-Fornells A. Human oscillatory activity associated to reward processing in a gambling task. Neuropsychologia 2008;46:241–248. [PubMed: 17804025]

Matsumoto M, Matsumoto K, Abe H, Tanaka K. Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci 2007;10:647–656. [PubMed: 17450137]

Miltner WHR, Braun CH, Coles MGH. 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;9:788–798.

Nieuwenhuis S, Holroyd CB, Mol N, Coles MG. Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. Neurosci Biobehav Rev 2004a;28:441–448. [PubMed: 15289008]

Nieuwenhuis S, Ridderinkhof KR, Talsma D, Coles MG, Holroyd CB, Kok A, van der Molen MW. A computational account of altered error processing in older age: dopamine and the error-related negativity. Cogn Affect Behav Neurosci 2002;2:19–36. [PubMed: 12452582]

Nieuwenhuis S, Slagter HA, von Geusau NJ, Heslenfeld DJ, Holroyd CB. Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. Eur J Neurosci 2005;21:3161–3168. [PubMed: 15978024]

Nieuwenhuis S, Yeung N, Holroyd CB, Schurger A, Cohen JD. Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. Cereb Cortex 2004b; 14:741–747. [PubMed: 15054053]

Oya H, Adolphs R, Kawasaki H, Bechara A, Damasio A, Howard MA 3rd. Electrophysiological correlates of reward prediction error recorded in the human prefrontal cortex. Proc Natl Acad Sci U S A 2005;102:8351–8356. [PubMed: 15928095]

Paus T. Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat Rev Neurosci 2001;2:417–424. [PubMed: 11389475]

Potts GF, Martin LE, Burton P, Montague PR. When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. J Cogn Neurosci 2006;18:1112–1119. [PubMed: 16839285]

Ridderinkhof KR, Nieuwenhuis S, Bashore TR. Errors are foreshadowed in brain potentials associated with action monitoring in cingulate cortex in humans. Neurosci Lett 2003;348:1–4. [PubMed: 12893411]

Ritter P, Becker R. Detecting alpha rhythm phase reset by phase sorting: caveats to consider. Neuroimage 2009;47:1–4. [PubMed: 19376248]

Sauseng P, Klimesch W, Gruber WR, Hanslmayr S, Freunberger R, Doppelmayr M. Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. Neuroscience 2007;146:1435–1444. [PubMed: 17459593]

Schultz W. Getting formal with dopamine and reward. Neuron 2002;36:241–263. [PubMed: 12383780]

Shima K, Tanji J. Role for cingulate motor area cells in voluntary movement selection based on reward. Science 1998;282:1335–1338. [PubMed: 9812901]

Sutton, RS.; Barto, AG. Reinforcement learning : an introduction. MIT Press; Cambridge, Mass: 1998.

Trujillo LT, Allen JJ. Theta EEG dynamics of the error-related negativity. Clin Neurophysiol 2007;118:645–668. [PubMed: 17223380]

Tsujimoto T, Shimazu H, Isomura Y. Direct recording of theta oscillations in primate prefrontal and anterior cingulate cortices. J Neurophysiol 2006;95:2987–3000. [PubMed: 16467430]

van Schie HT, Mars RB, Coles MG, Bekkering H. Modulation of activity in medial frontal and motor cortices during error observation. Nat Neurosci 2004;7:549–554. [PubMed: 15107858]

van Veen V, Holroyd CB, Cohen JD, Stenger VA, Carter CS. Errors without conflict: implications for performance monitoring theories of anterior cingulate cortex. Brain Cogn 2004;56:267–276. [PubMed: 15518940]

Vidal F, Burle B, Bonnet M, Grapperon J, Hasbroucq T. Error negativity on correct trials: a reexamination of available data. Biol Psychol 2003;64:265–282. [PubMed: 14630407]

Wang C, Ulbert I, Schomer DL, Marinkovic K, Halgren E. Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. J Neurosci 2005;25:604–613. [PubMed: 15659596]

Watkins CJCHD,P. Technical Note: Q-Learning. Machine Learning 1992;8:279.

Womelsdorf T, Schoffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P. Modulation of neuronal interactions through neuronal synchronization. Science 2007;316:1609–1612. [PubMed: 17569862]

Wrase J, Kahnt T, Schlagenhauf F, Beck A, Cohen MX, Knutson B, Heinz A. Different neural systems adjust motor behavior in response to reward and punishment. Neuroimage 2007;36:1253–1262. [PubMed: 17521924]

Yasuda A, Sato A, Miyawaki K, Kumano H, Kuboki T. Error-related negativity reflects detection of negative reward prediction error. Neuroreport 2004;15:2561–2565. [PubMed: 15538196]

Yeung N, Bogacz R, Holroyd CB, Cohen JD. Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. Psychophysiology 2004;41:822–832. [PubMed: 15563335]

Yeung N, Bogacz R, Holroyd CB, Nieuwenhuis S, Cohen JD. Theta phase resetting and the error-related negativity. Psychophysiology 2007;44:39–49. [PubMed: 17241139]

Yeung N, Holroyd CB, Cohen JD. ERP correlates of feedback and reward processing in the presence and absence of response choice. Cereb Cortex 2005;15:535–544. [PubMed: 15319308]

Yeung N, Sanfey AG. Independent coding of reward magnitude and valence in the human brain. J Neurosci 2004;24:6258–6264. [PubMed: 15254080]

A)

$$\text{Estimated Action Value} = \text{Last Action Value} + \text{Learning Rate}\,[\text{Prediction Error}]$$

$$Q\,(t+1) = Q\,(t) + \alpha\,[\text{reward}\,(t) - Q\,(t)]$$

Correct $\quad Q_A\,(t+1) = .5 + \alpha\,[1-.5]$

Incorrect $\quad Q_C = ...$

Correct $\quad Q_F = ...$

Incorrect $\quad Q_A\,(t+1) = .6 + \alpha\,[0 - .6]$

B)

Incorrect ——— Correct – – – Difference — —

1.

Q learning and the FRN. A) Example Q-learning algorithm and reinforcement learning task symbols (green boxes are not shown to participants). For example, if a participant picks stimulus 'A' and is rewarded, a positive prediction error will occur in relation to the extant action value. This value will update, increasing the likelihood of choosing 'A' in the future. The next time the A/B pair is shown, if the participant chooses 'A' again and is punished, the negative prediction error occurs in relation to the action value (which was large in this case, resulting in a more negative prediction error). B) Negative prediction errors have been proposed to be reflected by the Feedback Related Negativity (FRN) component of the ERP. Feedback locked current source density ERPs show correct, incorrect, and difference (incorrect minus correct) waves from the FCz electrode. The topographic map shows the power of the difference wave, averaged in window specified at the top of the graph (200-350 ms), demonstrating the timing and topography of the FRN.
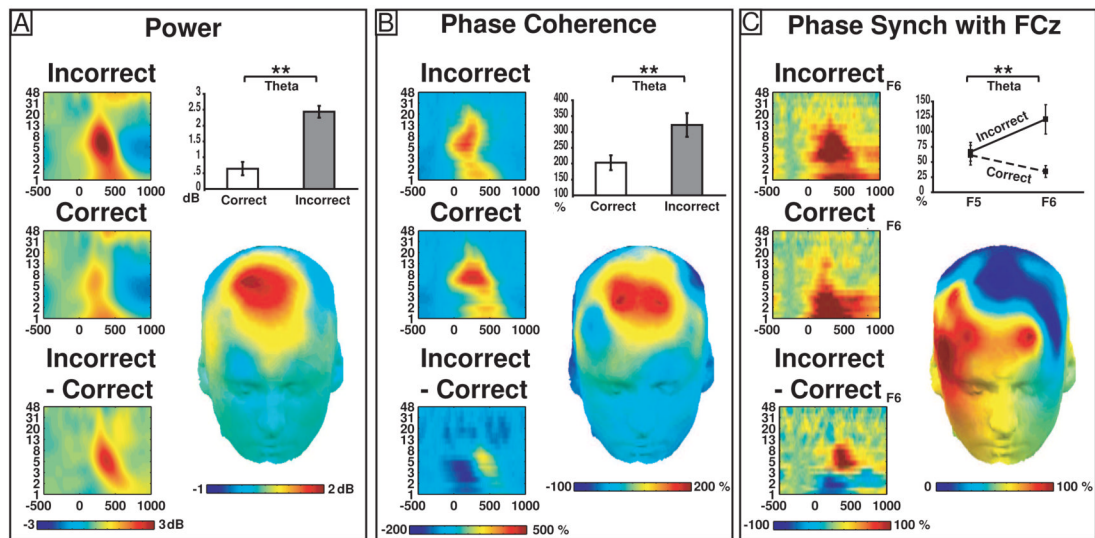
**2.**

Grand average time- frequency plots show correct, incorrect, and difference conditions for: A) CSD power changes (FCz), B) phase coherence increases from baseline (FCz), and C) phase synchrony increases from baseline (F6-FCz). A strong theta band increase in power, phase coherence, and phase synchrony can be seen following incorrect feedback ~ 300-500 ms; these effects are significantly different between conditions (as shown in the bar and line charts). Topographic plots show difference wave distributions (averaged over 300-500 ms) for the theta band.
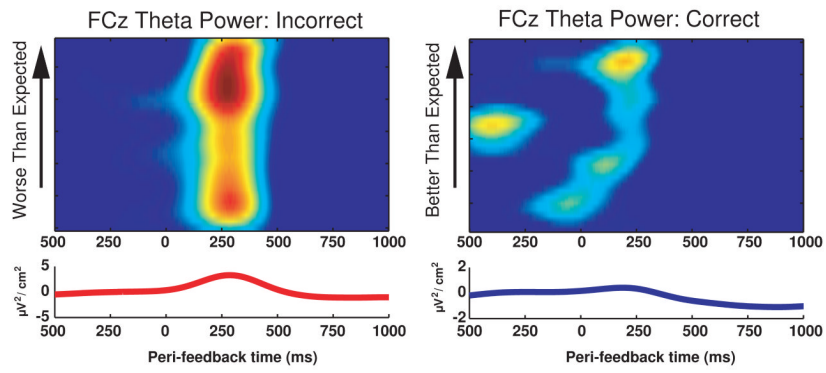
3.
Grand averaged theta CSD power ERPplots from the FCz electrode for incorrect and correct conditions, smoothed and sorted by degree of prediction error.
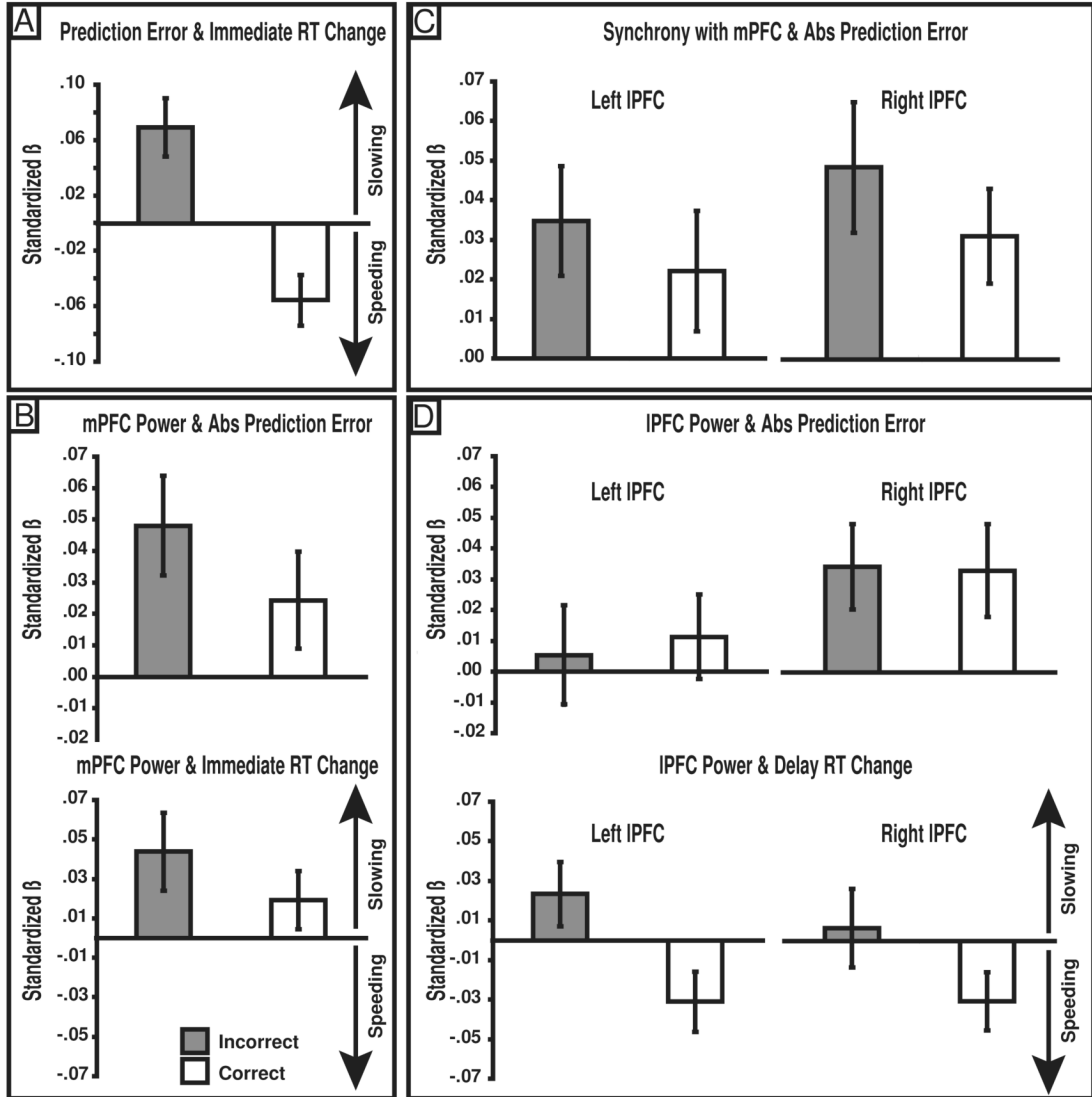
4.

Standardized beta weights (+/− SE) reflecting single-trial relationships as revealed by mixed model statistical tests. Note that prediction error is coded as an absolute measure for display purposes. A) The magnitude of prediction error is related to immediate reaction time slowing following incorrect feedback and speeding following correct feedback. B) mPFC theta power is related to the magnitude of prediction error, and mPFC power predicts immediate reaction time slowing on the next trial. C) Bi-lateral synchrony between mPFC and lPFC sites is related to the magnitude of prediction error. D) Right lPFC power is related to the magnitude of prediction error, and bilateral PFC power predicts reaction time speeding on the next trial of the same type (working memory-related speeding).
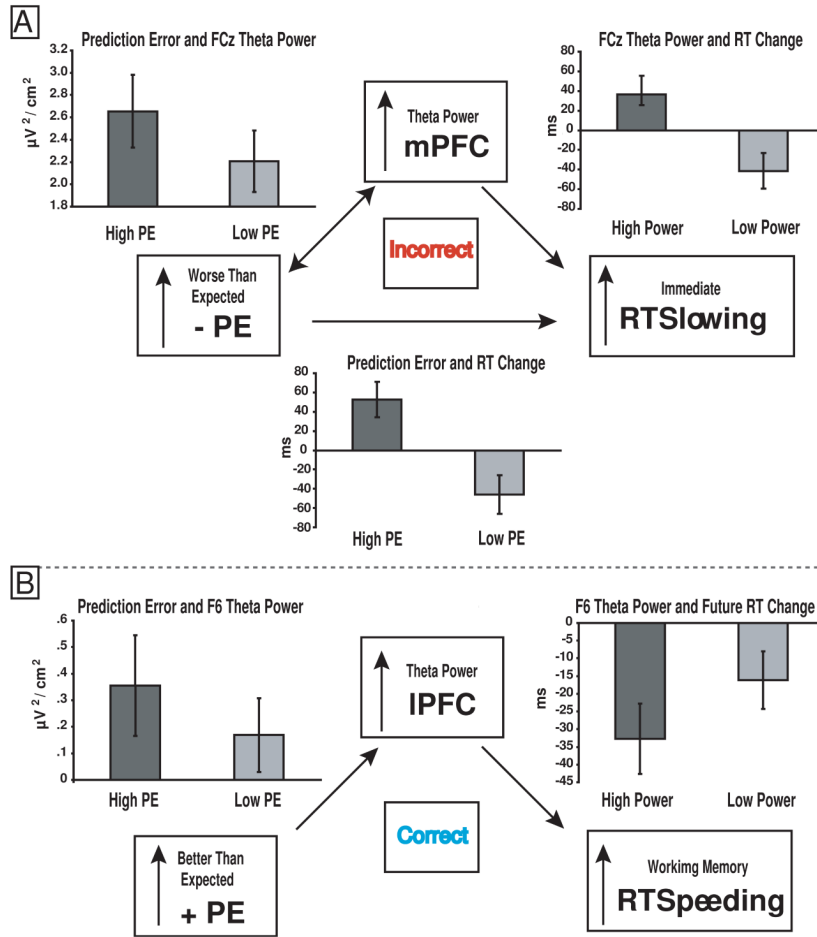
5.

Qualitative relationships between predictor variables (abcissa: median split) and the value of dependent variable (ordinate). A) Following incorrect feedback, the magnitude of negative prediction error and the amplitude of mPFC theta were directly related to each other. Both the magnitude of negative prediction error and the amplitude of mPFC theta predicted the degree of immediate reaction time slowing. Medial PFC theta power may be a reflection of a system that uses negative prediction errors to immediately adapt behavior. B) Following correct feedback, the magnitude of positive prediction error was directly related to the amplitude of lPFC theta power. Lateral PFC theta power predicted reaction time speeding for the same trial type the next time it was encountered (after a delay). Lateral PFC theta power may be a reflection of a system that updates working memory for stimulus value in the service of future behavioral adaptation.