Frontal theta links prediction errors to behavioral adaptation in reinforcement learning

James F. Cavanagh [a,*], Michael J. Frank [b], Theresa J. Klein [a], John J.B. Allen [a]

[a] Department of Psychology, University of Arizona, 1503 University Ave, Tucson, AZ, USA
[b] Brown University, Providence, RI, USA

ARTICLE INFO

ABSTRACT

Investigations into action monitoring have consistently detailed a frontocentral 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. Mediofrontal 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.

© 2009 Elsevier Inc. All rights reserved.

Investigations into action monitoring have consistently detailed a frontocentral 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 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 mediofrontal 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 errors 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

* Corresponding author.
  E-mail address: jim.f.cav@gmail.com (J.F. Cavanagh).

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, 2003; Nieuwenhuis et al., 2004a,b).

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 nonexistent (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 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, 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 feedbacks 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, 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, 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 to allow 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 feedbacks.