Cerebral Cortex November 2012;22:2575–2586
doi:10.1093/cercor/bhr332
Advance Access publication November 25, 2011

Frontal Theta Reflects Uncertainty and Unexpectedness during Exploration and Exploitation

James F. Cavanagh[1], Christina M. Figueroa[1], Michael X Cohen[2,3] and Michael J. Frank[1,4,5]

[1]Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI 02912, USA, [2]Department of Psychology, University of Amsterdam, 85715 Amsterdam, the Netherlands, [3]Department of Physiology, University of Arizona, Tucson, AZ 1018 WB, USA and [4]Department of Psychiatry 02912 and [5]Brown Institute for Brain Science, Brown University, Providence, RI 02912, USA

Address correspondence to James F. Cavanagh, Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Box 1821, Providence, RI 02906, USA. Email: jim.f.cav@gmail.com.

In order to understand the exploitation/exploration trade-off in reinforcement learning, previous theoretical and empirical accounts have suggested that increased uncertainty may precede the decision to explore an alternative option. To date, the neural mechanisms that support the strategic application of uncertainty-driven exploration remain underspecified. In this study, electroencephalography (EEG) was used to assess trial-to-trial dynamics relevant to exploration and exploitation. Theta-band activities over middle and lateral frontal areas have previously been implicated in EEG studies of reinforcement learning and strategic control. It was hypothesized that these areas may interact during top-down strategic behavioral control involved in exploratory choices. Here, we used a dynamic reward–learning task and an associated mathematical model that predicted individual response times. This reinforcement-learning model generated value-based prediction errors and trial-by-trial estimates of exploration as a function of uncertainty. Mid-frontal theta power correlated with unsigned prediction error, although negative prediction errors had greater power overall. Trial-to-trial variations in response-locked frontal theta were linearly related to relative uncertainty and were larger in individuals who used uncertainty to guide exploration. This finding suggests that theta-band activities reflect prefrontal-directed strategic control during exploratory choices.

Keywords: EEG, exploration, prediction error, reinforcement learning, uncertainty

Introduction

Goal-directed decision making involves not only making better decisions (exploiting) but also selecting alternative options when uncertain of their value (exploring). However, the circumstances that drive exploration remain undetermined (Cohen et al. 2007). Experiments have demonstrated that this process is not simply random: exploration may occur when long-term utility is low, when local costs are high, when the world has changed, or when there is uncertainty about alternative options (Daw et al. 2006; Cohen et al. 2007; Frank et al. 2009). Theoretical work has shown that reward can be maximized by increasing exploration when one is uncertain about reward statistics of alternative options (Gittins and Jones 1974; Dayan and Sejnowski 1996). However, exploration may be associated with cost due to forgone exploitation and an increased outcome risk (Cohen et al. 2007). It is clear that uncertainty-driven exploration is a potentially important facet of decision making, yet there is a dearth of empirical studies investigating the effects of uncertainty on exploration in humans. Moreover, there are currently no studies showing neural indices of uncertainty-driven exploration. Here we provide evidence that frontal theta-band oscillations reflect neurophysiological processes linking uncertainty and the exploration/exploitation trade-off to reinforcement learning.

Previous work has identified frontopolar areas, particularly on the right side, as well as bilateral intraparietal sulci that were associated with the selection of low-value decisions, which the authors defined as exploratory (Daw et al. 2006). Frontopolar areas are suggested to reflect a high-level system in the hierarchy of behavioral control (Koechlin and Summerfield 2007; Badre 2008), thus this finding has been interpreted as evidence for a top-down influence underlying the choice to explore. However, the models and neural signals in this study (Daw et al. 2006) did not provide evidence of guided exploration as a function of uncertainty.

In contrast, our previous investigation of exploration used a dynamic reward–learning task that tracked strategic adjustments and estimations of success, demonstrating that relative uncertainty predicted exploration (Frank et al. 2009). This relative uncertainty measure captured variance associated with responses outside of the current exploitative pattern. This effect was presumed to reflect prefrontally mediated strategic sampling of the reward structure, although this hypothesis has not yet been supported with recordings of neural activities. Since goal-directed decision uncertainty has been shown to correlate with bilateral frontopolar cortex activity (Yoshida and Ishii 2006), and this same brain area has been previously associated with exploration (Daw et al. 2006), it is possible that frontopolar activities in this dynamic reward–learning task might reflect aspects of uncertainty-driven exploration.

The temporal specificity of electroencephalography (EEG) provides a compelling methodological advantage for assessing trial-to-trial effects reflective of exploration and exploitation. EEG investigations of reinforcement learning have primarily assessed the mid-frontal component known as the feedback-related negativity (FRN), which reflects phase-locked theta-band activities thought to originate from the mid and posterior cingulate cortices (Miltner et al. 1997; Holroyd and Coles 2002; Luu et al. 2003). This FRN/theta signal has been proposed to reflect the calculation of a negative prediction error (Holroyd and Coles 2002), which is critical for learning how to exploit. This idea has been supported by recent investigations (Cavanagh et al. 2010; Chase et al. 2010; Ichikawa et al. 2010; Philiastides et al. 2010), yet others have questioned the valence specificity of this signal (Oliveira et al. 2007; Baker and Holroyd 2011).

It has been shown that medial and lateral frontal areas interact via theta-band phase dynamics during strategic control

(Hanslmayr et al. 2008; Cavanagh et al. 2009; Cohen and Cavanagh 2011) and during reinforcement learning (Cavanagh et al. 2010). Therefore, theta-band activities over lateral and frontal polar areas may be a candidate for identifying neural activities involved in the instantiation of strategic control for uncertainty-driven exploration. In sum, it was hypothesized that theta-band activities in distinct neural systems would reflect interactive processes when learning to exploit (medial) and deciding to explore (lateral/polar).

Materials and Methods

Subjects

Seventeen (5 male) participants were recruited through the undergraduate subject pool at the University of Arizona. All participants gave informed consent for the project approved by the University of Arizona Research Protections Office. Participants ranged in ages from 17 to 21 years (mean = 18.8 years).

Task

This task has previously been used to investigate trial-specific reinforcement learning and exploration in genetic, patient, and pharmacological cases (Moustafa et al. 2008; Frank et al. 2009; Strauss et al. 2011). Participants were presented the outline of a circle; a small ball traced the outline of the circle (0–360°) over the course of 4 s (Fig. 1*a*). Instructions were as follows:

You will see a circle. A ball will make a full turn around the circle in 4 seconds. To win points, you need to stop the ball somewhere along this circle. To stop the ball, press the spacebar with your dominant index finger.

The time at which you respond affects in some way the number of points that you can win. Sometimes you will win lots of points and sometimes you will win fewer. Try to win as many points as you can! (If you don't respond by the end of the ball's cycle, you will not win any points). Hint: Try to respond at different times along the ball's cycle in order to learn how to win the most points.

The trial ended after the participant made a response or if the 4-s duration elapsed without a response. Probabilistic feedback was then presented 500 ms after the response for 1000-ms duration. Response time (RT) was used to calculate the probability of winning points on that trial and what the magnitude of points was (described below). Following feedback, there was an intertrial interval that was randomly calculated to be between 500 and 1500 s. Participants were also warned:

The length of the experiment is constant and is not affected by when you respond. In other words, responding quickly won't get you out of here faster!

There were 4 different blocks with 120 trials each; each block consisted of 1 of 4 counterbalanced conditions with differing RT-determined reward contingencies (Fig. 1*b–d*). Information on the different block types was also presented in the instructions:

In addition, this experiment also consists of four independent blocks. These blocks are marked by different colored circles. You should attempt to win the most points in each block. Try to respond at different times along the circle to learn how to win the most points in each block.

In 3 of these conditions, the magnitude of reward increased over the duration of the trial (4 s), yet the probability of reward decreased, as shown in Figure 1*b,c*. The reward functions were calculated so that expected value (EV: probability × magnitude) increased over time in one condition (IEV: slower responses were better), decreased over time in one condition (DEV: faster responses were better), or stayed constant over time (CEV: all responses were equal in EV) (see Fig. 1*d*). This CEV condition was used as a control condition. The fourth condition had an increasing probability yet a decreasing magnitude of reward (also with constant EV over the duration of the trial), providing a reversed CEV condition (CEV-R: all responses were equal in EV). Compared with CEV, slower RTs in the CEV-R condition are interpreted as reflecting a risk aversive performance strategy in that a high probability of reward is preferred over a high magnitude of reward (this has been commonly observed in the aforementioned previous applications of this task). While performance on each block was separately analyzed to ensure that participants learned the task, trials from all conditions were aggregated for EEG analyses.

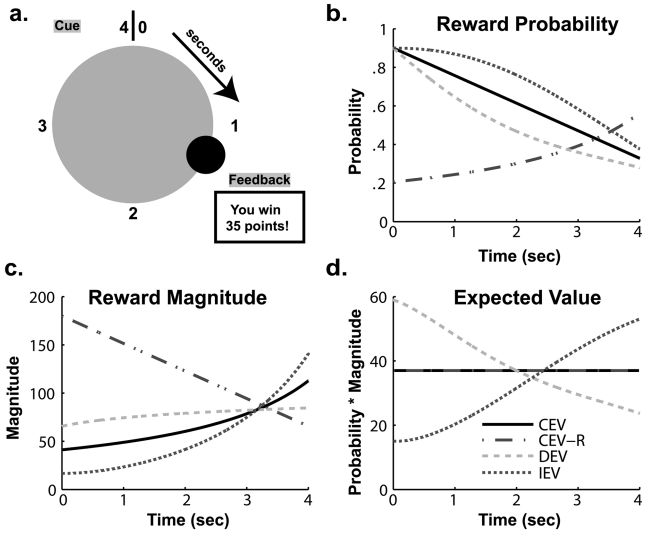

Figure 1. The task consisted of 4 blocks with varying EV of reward depending on RT. (*a*) Example trial where the participant made a response ~1.2 s, receiving a 35-point reward. Blocks consisted of different conditions that had varied probabilities (*b*) and magnitudes (*c*) or reward over time, creating separate EV distributions (*d*). To maximize reward, participants need to learn to respond faster when EV decreased (DEV) and slower when EV increased (IEV) compared with when it was constant (CEV). A fourth condition contained reversed magnitudes and probabilities with constant EV (CEV-R); slower responses on this block reflect a preference for magnitudes over probabilities and a risk-averse strategy.

Algorithmic Model of Performance

The computational model was applied to each block of the data in order to capture variance associated with strategic trial-to-trial adaptation (described in Frank et al. 2009). Participants were assumed to update an expected reward value (V) after each feedback at time t with fixed learning rate $\alpha = 0.1$:

$$V(t+1) = V(t) + \alpha(\delta),$$

where the update is based on reward prediction errors (δ):

$$\delta = [\text{Reward}(t) - V(t)].$$

These reward prediction errors can be positive or negative if the reward was better or worse, respectively, than the current value estimate V. To capture strategic processes, probability density functions (PDFs) were used to reflect participant's belief distributions that specific actions would yield a better than average outcome (i.e., a positive prediction error; see Fig. 2*a*). Since reward functions are monotonic, one does not have to track reward statistics for all possible RTs. Rather, one only needs to track reward statistics for "slower" or "faster" responses (e.g., as compared to the ongoing average RT) and then adjust RTs in proportion to (and in the direction of) their difference. Thus, when faster RTs lead to reliably better outcomes than slower RTs, this model predicts that participants would respond proportionally faster. The model also predicts that subjects would explore other responses when they are particularly uncertain about their reward statistics, where uncertainty is quantified by the variance in the belief distributions, as described below.

The belief distributions about reward statistics for slower or faster responses were updated after each trial via Bayes' rule. We use beta distributions to represent these beliefs, in which Bayesian updating simply amounts to incrementing hyperparameters η and β with each positive and negative prediction error, respectively. The resulting PDF is shown in Figure 2*a* and is as follows:

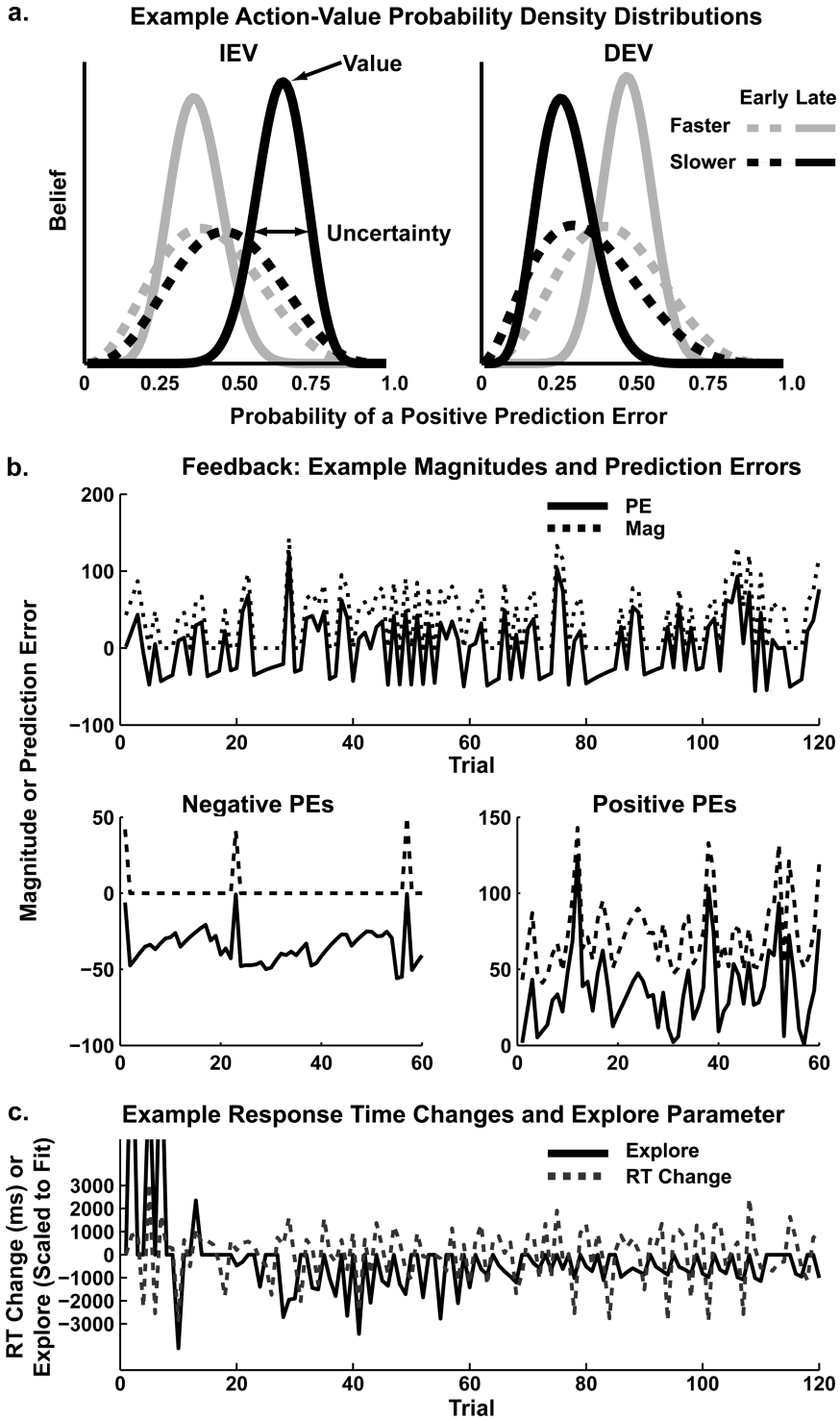

Figure 2. Example model outcomes from a single representative participant. (*a*) PDFs. Early in the block, probability densities peak around 0.5 and have large variances. Later in learning, these migrate and sharpen to provide estimates of value (mean) and uncertainty (variance). This investigation used the relative difference between slower and faster variances to classify exploitative and exploratory strategies. (*b*) Feedback and prediction error. While positive prediction errors closely followed the magnitude of reward, negative prediction errors largely occur to zero-reward feedback and were estimated from the model. (*c*) Explore parameter and RT change. Large differences in uncertainty predicted RT swings in participants with high explore parameters.

$$f\left(x;\eta,\beta\right)=\frac{X^{(\eta-1)}\times(1-X)^{\beta-1}}{\int_0^1 Z^{(\eta-1)}\times(1-Z)^{\beta-1}\,\mathrm{d}z},$$

where the denominator reflects the beta function.

The probabilistic EV and uncertainty for each type of response (slower or faster) was then computed analytically as the mean $\mu = \eta/(\eta + \beta)$ and variance $\sigma^2 = (\eta \times \beta)/[(\eta + \beta)^2 \times (\eta + \beta + 1)]$ of the PDF. This measure of variance quantifies the uncertainty about the value of each response type (fast or slow). As participants make faster or slower responses, the means come to reflect the probability that a better than average outcome (positive prediction error) will be experienced for that response type. The variance (uncertainty) about

these means decrease with experience (albeit at a slower rate for more variable outcomes). RT adjustments are well captured in this task by assuming that participants adjust RTs in proportion to the relative difference between the means (μ) of slower and faster RTs [$\mu_{slow}(t) - \mu_{fast}(t)$], scaled by free parameter ρ (i.e., the degree to which individual subjects adjust RTs by EV). Because participants could not be confident about the value of the means without sufficient experience, we further assumed that participants might explore in proportion to the relative difference in uncertainty between faster and slower responses. Exploration was included in the model by adding the following term: [$\sigma_{slow}(t) - \sigma_{fast}(t)$], scaled by free parameter ε. If this fitted ε parameter was positive, participants were more likely to direct exploration toward the more uncertain response. Trials immediately following exploratory choices were forced to be 0 to avoid mis-estimation of exploration due to recent task dynamics (see methods in Frank et al. 2009; Fig. 2*c*).

The model used here was identical to that used previously, which provided the best fit to behavior (Frank et al. 2009; Strauss et al. 2011), and included other free parameters to improve model fit including baseline response speed with parameter *K*, autocorrelation of preceding RT scaled by parameter λ, a putative striatal tendency to generally speed up following positive prediction errors (α*G*) and slow down following negative prediction errors (α*N*), and a "going for the gold" factor that reflected seeking the highest rewarded RT thus far [$RT_{best} - RT_{avg}$], scaled by parameter ν. In sum, free parameters (*K*, ρ, ε, λ, α*G*, α*N*, *v*) were fit using a simplex method (Matlab function "fmincon") to minimize the sum of squared errors between predicted and actual RTs across all trials for each subject (see Table 1). The model took the form

$$RT(t) = K + \rho\left[\mu_{slow}(t) - \mu_{fast}(t)\right] + \varepsilon\left[\sigma_{slow(t)} - \sigma_{fast}(t)\right] + \lambda RT(t-1) - \sum \alpha G(\delta)_{+} + \sum \alpha N(\delta)_{-} + v\left[RT_{best} - RT_{avg}\right].$$

Note, however, that none of these additional parameters influence the estimates of prediction error, mean or uncertainty, and thus any neural response to these variables is independent of other aspects of the model. In other simulations, we have further confirmed that the model ranking of subjects' tendencies to guide exploration by uncertainty is robust to various assumptions about the other forms that the model might take.

Trial-to-trial prediction errors (δ) were used as predictor variables in regressions with feedback-locked EEG power (see Fig. 2*b*). As in the previous investigations (Frank et al. 2009; Strauss et al. 2011), the relative difference in uncertainties was used to predict exploratory decisions. Here the difference between the chosen and the unchosen response options [$\sigma_{chosen}(t) - \sigma_{unchosen}(t)$] was specifically investigated in the context of trial-to-trial event-related EEG power. For example, since responses were characterized as either slow or fast, in the case of RT slowing this chosen–unchosen difference would reflect the slow PDF variance minus the fast PDF variance (and the opposite for RT speeding). Therefore, a larger uncertainty value corresponded to greater relative uncertainty for the chosen response.

It is important to note that some studies have described how uncertainty "negatively" influences exploration due to ambiguity aversion (Payzan-LeNestour and Bossaerts 2011). A critical difference between the reinforcement-learning task used here and other tasks is the absence of structural uncertainty (the reward dynamics do not shift or reverse within a block) and the fact that here exploration can be used to "reduce" future uncertainty in the long run. These issues are discussed in greater depth in the discussion. In order to formally test for any bidirectional influence of uncertainty on RT, we ran another model in which we first defined exploratory trials (those selecting the action with lower EV; Daw et al. 2006) and then refit the uncertainty-exploration parameter "only" to those trials to determine whether exploratory trials in particular were more often driven toward the most uncertain option. This procedure prevented the fitting procedure from penalizing model fit in all the exploitation trials in which the more certain action was generally selected. Consequently, participants were categorized as using uncertainty to drive exploration (those with positive ε) or not (those with nonpositive ε) (see Table 1). This characterization also allowed us to investigate whether neural responses to uncertainty differed between those estimated to have positive ε and those that did not.

Table 1
Parameter estimations for both models, mean (SD)

Parameter	Frank et al. (2009) model	Bidirectional uncertainty
Baseline response speed *K*	1134 (347)	1196 (339)
Autocorrelation of RTs λ	0.32 (0.17)	0.30 (0.16)
Highest reward thus far *v*	0.10 (0.10)	0.12 (0.11)
Gain learning rate α*G*	0.11 (0.26)	0.29 (0.26)
Loss learning rate α*N*	0.13 (0.27)	0.29 (0.25)
Mean difference scaling ρ	620 (390)	473 (394)
Variance difference scaling ε	1937 (2437)	1027 (4529)
N subjects with positive ε/mean (SD)	13/2353 (2506)	9/4583.67 (2053)
N subjects with nonpositive ε/mean (SD)	4/0 (0)	8/-2974 (2742)

Electrophysiological Recording and Processing

Scalp voltage was measured using 58 Ag/AgCl electrodes, plus 2 mastoid sites, referenced to a site immediately posterior to Cz using a Synamps^2 system (band-pass filter 0.5–100 Hz, 500 Hz sampling rate, impedances <10 kΩ), re-referenced offline to averaged mastoids. User-identified bad epochs containing movement or muscle artifact, large voltage shifts, or amplifier saturation were marked and removed (mean = 5.6%, standard deviation [SD] = 3.3%), and bad channels were interpolated. An infomax independent components analysis was run for each subject using "runica" from the EEGLab toolbox; components associated with eye blinks were removed (Delorme and Makeig 2004).

Event-Related Potentials

Feedback-locked event-related potentials (ERPs) were split by sign (negative, positive) and size (big = above median, small = below median) of prediction errors. These ERPs were low-pass filtered at 20 Hz since most of the ERP variance associated with prediction errors are in low-frequency bands (Cavanagh et al. 2010; Chase et al. 2010; Ichikawa et al. 2010; Philiastides et al. 2010). For analytic purposes, consecutive peaks and valleys in the ERP were defined as P2 (176 ms), FRN (276 ms), P3 (376 ms), and N4 (476 ms). While these labels may not converge with some traditional definitions of ERP components, they represent a logical nomenclature for defining ERP morphology. For example, the term FRN is used instead of N2 to denote the specificity of this component to reinforcement feedback, although we acknowledge the ambiguity of this label given that numerous studies that have suggested these are nondistinct entities (Holroyd and Coles 2002; Holroyd et al. 2008; Cavanagh et al. forthcoming). Values were taken as the mean of the ERP in a ±50-ms window around the peak/trough time indicated above. Baseline-independent amplitudes were taken by subtracting the preceding or following negativity from the local positivity (P2-FRN, P3-FRN, P3-N4).

Time–Frequency Calculations

Time–frequency calculations were computed using custom-written Matlab routines (Cohen et al. 2008; Cavanagh et al. 2009). Time–frequency measures were computed by multiplying the fast Fourier transformed (FFT) power spectrum of single-trial EEG data with the FFT power spectrum of a set of complex Morlet wavelets and taking the inverse FFT. The wavelet family is defined as a set of Gaussian-windowed complex sine waves: $e^{-i2\pi tf}\, e^{-t^2/(2\times\sigma^2)}$, where *t* is time, *f* is frequency (which increased from 1 to 50 Hz in 50 logarithmically spaced steps), and σ defines the width (or "cycles") of each frequency band, set according to $4/(2\pi f)$. The end result of this process is identical to time-domain signal convolution, and it resulted in 1) estimates of instantaneous power (the magnitude of the analytic signal), defined as $Z[t]$ (power time series: $p(t) = \text{real}[z(t)]^2 + \text{imag}[z(t)]^2$), and 2) phase (the phase angle) defined as $\varphi_t = \arctan(\text{imag}[z(t)]/\text{real}[z(t)])$. Each epoch was then cut in length to remove edge artifacts. Power was normalized by conversion to a decibel scale (10 × log10[power(*t*)/power(baseline)]), allowing a direct comparison of effects across frequency bands. The baseline for each frequency consisted of the

average power from −250 to 0 ms prior to the onset of the cues. Whereas the ERPs reflect phase-locked EEG variance, these time–frequency measures reflect total power (phase locked and phase varying). See Figure 3 for plots of power at the mid-frontal FCz electrode, with overlapping ERPs and topographic plots of power in the time–frequency regions of interest (hereafter referred to as TF-ROIs) identified by dashed boxes.

Statistical Analyses

Analyses of prediction error aimed to separately assess the influence of sign and magnitude of feedback expectancy. Separate regressions were run for 1) all trials, 2) trials associated with a positive prediction error, and 3) trials associated with a negative prediction error, with an additional contrast of 4) negative–positive prediction errors. These analyses investigated the relationship between the absolute value of the prediction error and feedback-locked EEG power (across time and frequency points at the FCz electrode and across channels for the FCz-defined TF-ROI). Measures of regression slope (standardized β weight) and intercept were retained at each time–frequency point for each participant, providing separate estimates of EEG activities associated with prediction error sign (intercept) and magnitude (slope). Data points that had EEG activities 3 SDs beyond the mean were removed prior to regression. Figure 5 displays mean slopes and intercepts across participants. Intercepts were converted to decibel change from the pre-cue baseline intercept for display.

Analyses of uncertainty used Spearman's ρ correlations due to nonlinearity in the relative uncertainty regressor. These correlations were also calculated across time and frequency points at mid-frontal (Cz) and right frontopolar (FP2) electrodes. Additional contrasts used independent *t*-tests to compare the transformed Fisher's ρ-to-*z* coefficients of the 2 subgroups defined by the bivalenced uncertainty model (positive ε > nonpositive ε), in order to determine whether neural responses to uncertainty differed between these subgroups.

TF-ROIs were included given the strong a priori hypotheses about the role of frontal theta in feedback- and response-locked signals. TF-ROIs were defined based on the grand average time–frequency plots in Figure 3 (response: 3–4.5 Hz, 0–150 ms; feedback: 4–8 Hz, 250–450 ms). Figure 7 shows extended pre-response TF-ROIs (−250 to 150 ms) to detail the range of effects in relevant topoplots. All time–frequency correlation coefficient plots were thresholded by only including pixels that were significantly ($P < 0.05$) above zero (prediction error slope; uncertainty correlations), different from baseline (prediction error intercept), different between conditions (prediction error negative > positive contrasts), or different between groups (uncertainty positive ε > nonpositive ε) with a minimum cluster size of 200 voxels. In Figure 8, major findings are summarized by displaying the slope from the unthresholded TF-ROIs in Figures 5 and 7.

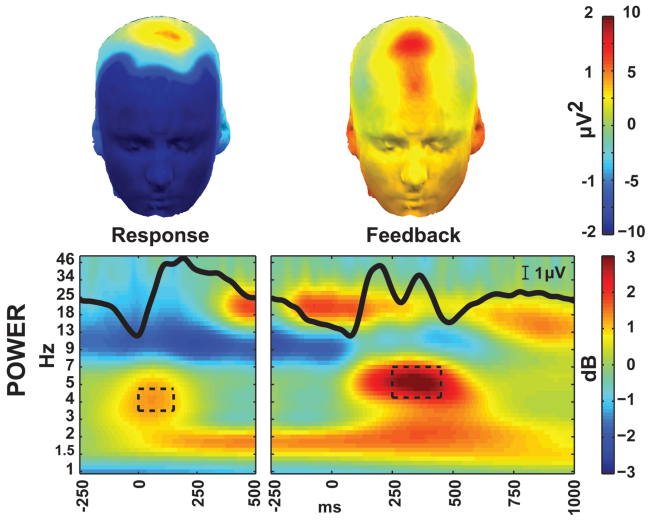

Figure 3. Event-related EEG at the FCz electrode to response and feedback. Positive amplitude is displayed up on the *y*-axis. The ERP is superimposed over the power plots in black. TF-ROIs are shown in the black boxes. Topomaps are taken from the TF-ROIs shown here (±10 μV^2 for feedback, ±2 μV^2 for response).

Results

Performance

As in the previous investigations with this task (Moustafa et al. 2008; Frank et al. 2009; Strauss et al. 2011), RTs were averaged within each block. Participants learned to exploit the task structure by reliably slowing down when EV increased and speeding up when EV decreased (IEV > CEV: $t_{16} = 3.95$, $P = 0.0012$; DEV > CEV: $t_{16} = -3.80$, $P = 0.0016$). Also replicating previous findings, participants displayed a relatively risk-averse strategy in the comparison between the 2 constant EV conditions: participants preferred high probabilities over high magnitudes of reward as evidenced by relative slowing in the CEV-R condition (CEV-R > CEV: $t_{16} = 2.38$, $P = 0.03$). See Figure 4*a*,*b*.

Model parameters were similar to our previous investigation, as shown in Table 1. Only 13 of the 17 participants had positive ε coefficients when this parameter was free to vary from zero to positive values (Frank et al. 2009; Strauss et al. 2011); thus, only these participants may be characterized as using uncertainty to drive exploration. Indeed, similar to the previous studies, there was a reliable correlation between the explore term based on relative uncertainty in a given trial and the RT adjustment from the previous trial in these 13 participants (single-trial regression weights mean = 0.22, SD = 0.13, 1-sample $t_{13} = 6.29$, $P < 0.001$). The additional bidirectional uncertainty model, which allowed this parameter to be fit with either positive, zero, or negative values, revealed that 9 of these participants still had positive ε coefficients while 8 participants did not (Table 1). These positive ε participants still had a reliable positive correlation between relative uncertainty and RT adjustment (mean = 0.20, SD = 0.21, 1-sample $t_8 = 2.88$, $P = 0.02$), whereas nonpositive ε participants did not (mean = 0.01, SD = 0.21, 1-sample $t < 1$). Although the difference between these groups was large, it was not statistically significant ($t_{15} = 1.91$, $P = 0.075$, see Fig. 4*c*). Nevertheless, at least 9 of the participants in this investigation can be quantified with this model as using uncertainty to drive exploration. Consequently, subsequent investigations of uncertainty were split between these positive ε and nonpositive ε subgroups.

Regression Coefficients: Prediction Error and Theta Power

Figure 5 shows the results of significant single-trial regressions between prediction error and feedback-locked EEG power. The same feedback-locked theta-band TF-ROI outlined in Figure 3 is detailed here. Time–frequency plots show the regression slope (standardized beta weight) or the intercept at each time–frequency point. The regression slopes demonstrate that there was a direct relationship between the absolute value of prediction error and theta power within this TF-ROI. This relationship did not differ between reward and punishment, as detailed by the difference plot in Figure 3. This null finding stands in contrast to the intercept, which demonstrated a significant increase for negative compared with positive prediction errors. Thus, the mere presence of a negative prediction error was associated with greater theta power

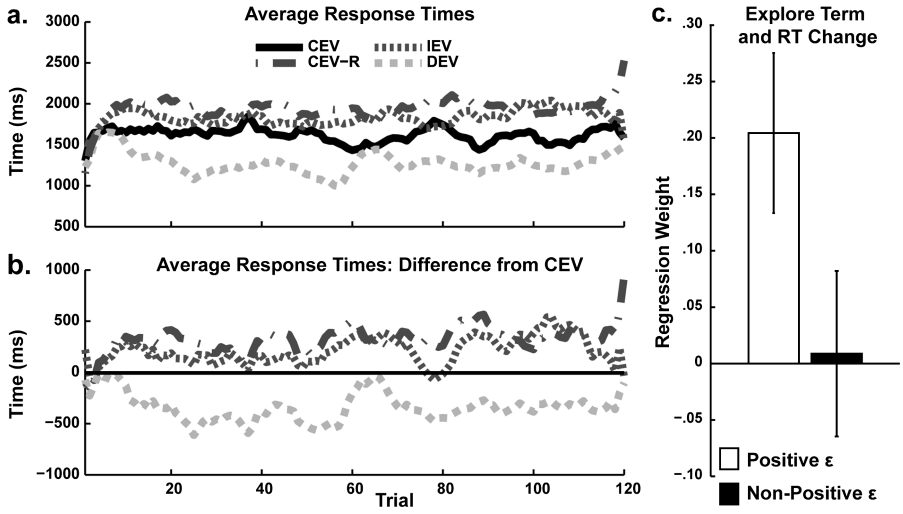

Figure 4. Average RT performance. (*a*, *b*) Participants learned to adapt RTs (smoothed with a 5-point kernel) depending on block-specific reward structure. For example, participants responded slower when EV increased over time and they responded faster when EV decreased over time, even relative to their own performance when EV was constant. Slower RTs to the constant EV-reversed (CEV-R) condition demonstrate a preference for reward probability over magnitude (risk aversion). (*c*) Positive ε participants were characterized by the model as using relative uncertainty to guide exploration. Here it can be seen that this group was characterized by significant correlations between trial-to-trial measures of exploration and RT change.

(compared with positive), but the degree to which EEG power scaled with greater deviations from expectation (larger prediction errors) was similar for both valences. Other time and frequency areas were different between valences, particularly in early delta and late theta through beta bands. These mid-frontal EEG findings in the theta-band TF-ROI support prior interpretations of a binary sign difference and a continuous magnitude effect in dorsal cingulate coding of reinforcement feedback (Hayden et al. 2011).

Comparison of FRN and Theta Power Split by Prediction Error

To simultaneously visualize these separate sign-related offset and magnitude-related scaling effects in formats common to EEG analyses, ERPs and theta power were binned by small and large valenced prediction errors (Fig. 6). Given the clear hypotheses about these effects and the descriptive nature of this analysis, separate general linear models were simply calculated for each of the time bins for each EEG feature.

For FRN amplitudes in the P2-FRN range, there was only a significant main effect for sign ($F_{1,16} = 5.32$, $P = 0.035$). In the P3-FRN time range, there was a main effect for sign ($F_{1,16} = 15.19$, $P < 0.001$), a main effect for magnitude ($F_{1,16} = 12.25$, $P = 0.003$), and an interaction ($F_{1,16} = 6.16$, $P = 0.025$), revealing significant contrasts between big ($P = 0.002$), small ($P = 0.007$), and negative ($P = 0.001$) prediction errors. In the P3-N4 time range, there was again only a main effect for sign ($F_{1,16} = 8.24$, $P = 0.01$). In sum, 4 of 5 significant main effects or simple contrasts revealed bigger amplitudes for negative prediction errors. Only a single statistical test revealed a within-sign effect of magnitude, fitting with a signed prediction error account: P3-FRN big > small negative prediction error amplitude (Fig. 6*a*). However, task dynamics may have contributed to an imbalanced consistency between these conditions: 20% of small negative trials consisted of nonzero rewards compared with less than 1% for big negative trials.

For theta power (Fig. 6*b*), there were only main effects of sign: in the P2-FRN time range ($F_{1,16} = 8.67$, $P < 0.01$), in the P3-FRN time range ($F_{1,16} = 18.82$, $P < 0.001$), and the P3-N4 time range ($F_{1,16} = 21.39$, $P < 0.001$). This overall pattern fits with the FRN findings describing a general effect of larger EEG power for negative prediction errors, and it strongly matches the findings of the regression weights detailing a sign-related intercept offset for worse-than-expected outcomes.

Uncertainty

Figure 7 details the trial-to-trial correlations between EEG power and the relative uncertainty of the chosen minus the unchosen option (Spearman's ρ values). The leftmost time–frequency plots include all $N = 17$ participants (the $N = 13$ participants with positive explore parameters in the model of Frank et al. (2009) revealed the same significant patterns). To demonstrate that these uncertainty–EEG correlations are related to exploration, the rightmost time–frequency plots show the significant statistical differences in the difference between ρ correlation coefficients between participants who used uncertainty for exploration (positive ε) and those who did not (nonpositive ε) as defined by the bivalenced exploration model.

Total relative uncertainty was correlated with mid-frontal beta power and middle/frontopolar delta/theta power in the time period around response commission (Fig. 7*a*). These uncertainty-related responses were only present when participants actually selected the uncertain option (exploratory trials). Figure 7*b* shows how increasing uncertainty correlated with greater EEG power in the delta/theta range in both right lateral/frontopolar and mid-frontal areas during exploration, preceding the response by up to 500 ms. These theta-band effects were significantly greater in positive ε compared with nonpositive ε participants, although the frontopolar effect differentiating these groups occurred at somewhat earlier time points and at a lower frequency than was expected. In contrast, Figure 7*c* shows how exploitative responses were characterized by negative relationships between the degree of relative uncertainty and a broad topography and spectra of EEG power. This negative relationship occurs in exploitation because as

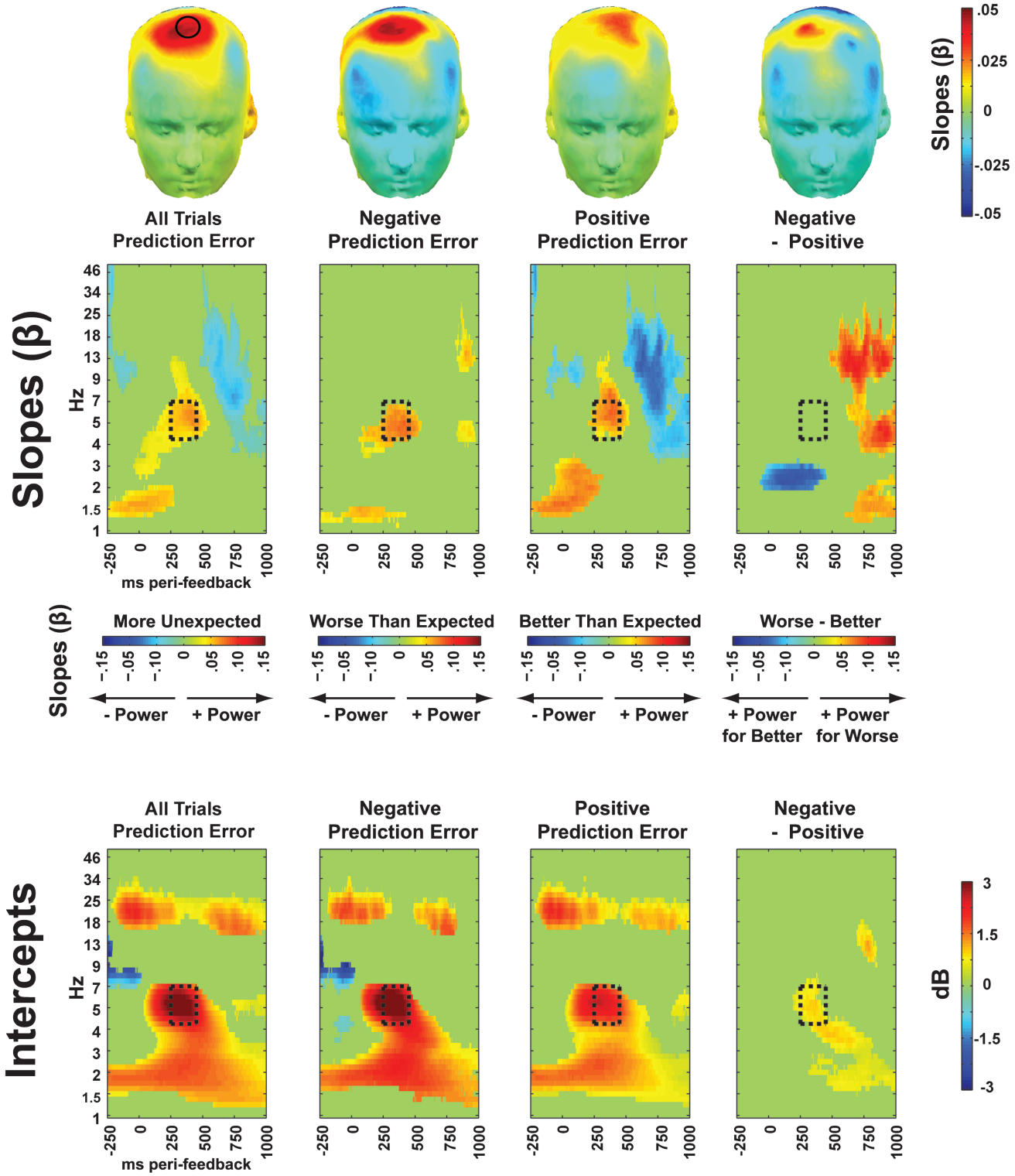

Figure 5. Mid-frontal (FCz) EEG relationships with prediction error measures. Topographic plots show regression slopes in the TF-ROIs. FCz is indicated on the first topoplot. Time–frequency plots of slopes show the regression weight with absolute prediction error, demonstrating that prediction error magnitude scales with mid-frontal theta. This theta-band relationship does not depend on the sign of the prediction error, as demonstrated by a lack of statistical difference between negative and positive prediction errors in the theta-band TF-ROI. In contrast, there was an intercept offset for negative compared with positive prediction errors in this TF-ROI (converted to decibel change from pre-cue baseline intercept). These plots demonstrate how mid-frontal theta does not reflect a signed prediction error; rather, it reflects the overall degree of surprise with an asymmetrical offset when events are worse than expected.

subjects choose the increasingly more certain (exploitative) action, there is lower EEG power. There were no significant differences between ε groups in this exploit condition.

Together, these data suggest that frontal theta linearly scales with the relative uncertainty of the chosen option. When peri-response theta power over these sites was high, individuals

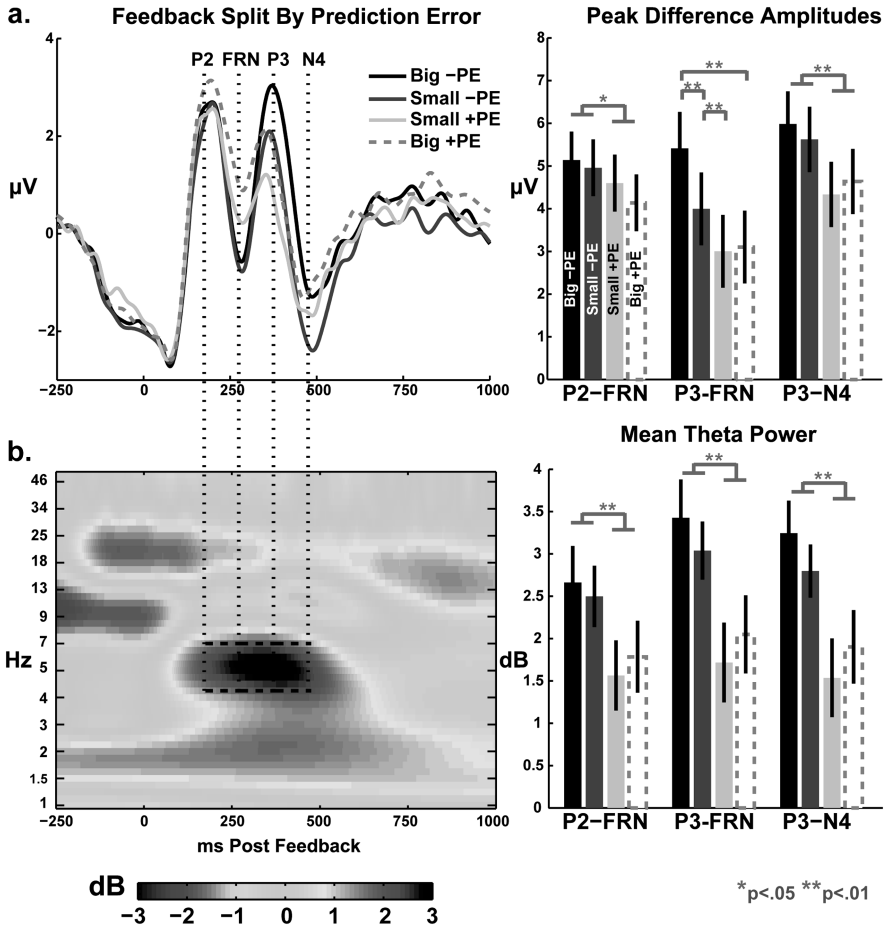

Figure 6. Feedback-locked ERP (*a*) and total theta power (*b*) from the FCz electrode, split by sign and magnitude of prediction error. Bar graphs display mean values (±SEM) taken from the time windows indicated by vertical dashed lines. For both EEG measurements, there was a strong and reliable effect for sign (negative > positive prediction error).

were more likely to explore the uncertain option in that trial. This theta-uncertainty relationship was greater in those subjects who were more sensitive to uncertainty in their decisions to explore. These findings are consistent with a hypothesis for prefrontally directed strategic control during exploratory choices.

Summary

Figure 8 summarizes the major methods and findings, including "when" different effects occur, "how" conditions were defined with model outcomes, "where" the effects are located on the scalp, and "what" the patterns of correlation were for different variables. Note that the exploitation slope is inverted in this figure compared with Figure 7*c* in order to detail a single continuum of theta power correlation with relative uncertainty for the selected response.

Discussion

This investigation revealed that distinct frontal EEG responses, primarily in the theta band, reflect neural measures related to unexpectedness updating and uncertainty-driven exploration. Advancing previous findings, here we characterized how the medial theta-band response to feedback does not reflect a negative reward (punishment) prediction error; rather it reflects an asymmetrical sensitivity to worse-than-expected events while directly scaling with the degree of unexpectedness. We found novel evidence that information related to uncertainty is encoded during the decision time surrounding response commission. Notably, enhanced theta power related to uncertainty was specifically observed in trials in which participants chose the uncertain option and primarily in participants who used uncertainty to explore. This finding provides novel evidence for a link between frontal cortical activities and uncertainty-driven exploration.

Unexpectedness and Mid-Frontal Theta

Both mid-frontal theta and the mid-cingulate cortex are particularly reactive to signals of novelty, error, punishment, and conflict, yet also show a diminished response to reward (Shima and Tanji 1998; Bush et al. 2002; Luu et al. 2003; Wang et al. 2005). Given the role of mid-frontal theta and the mid-cingulate cortex in evaluating performance and adjusting behavior (Debener et al. 2005; Kennerley et al. 2006; Behrens et al. 2007), error and punishment signals may be reliable indicators of the need to adjust behavior—whereas correct responses and rewards usually indicate that performance is on the right track. However, it is likely that the dynamic reward optimization task used here revealed a more sensitive role of mediofrontal systems in unexpectedness updating and behavioral adaptation (i.e., not specific to negative prediction errors).

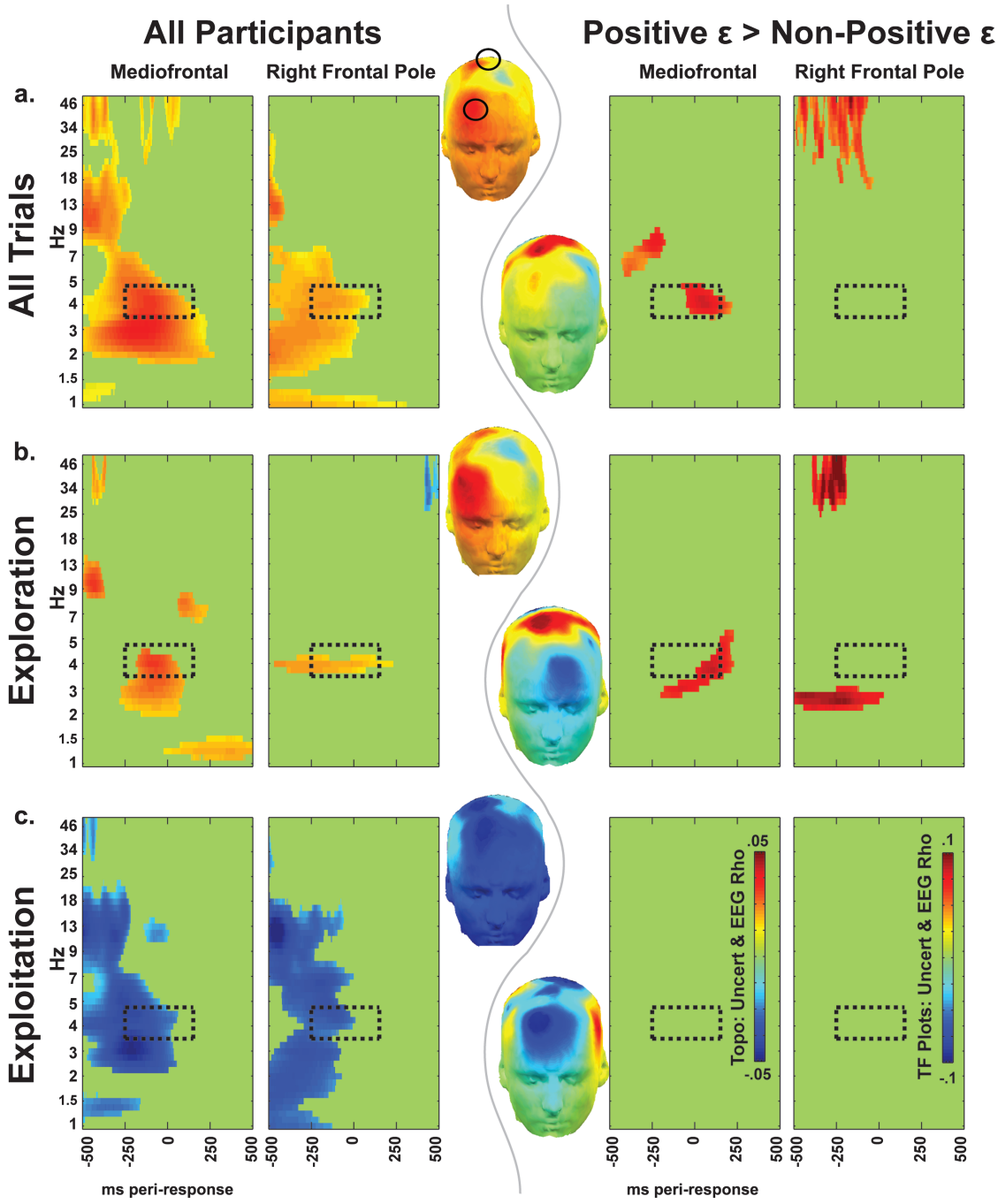

Figure 7. Nonparametric correlations (ρ values) between relative uncertainty (chosen–unchosen option) and EEG power at the time of the response, for all subjects (left columns) and positive ε > nonpositive ε groups (right columns). Topographic plots show EEG–uncertainty correlations in the extended TF-ROI (leftward topoplots are all subjects, rightward topoplots are positive ε > nonpositive ε). Mid-frontal and right frontopolar sites are indicated on the top topoplot. (*a*) Total relative uncertainty is characterized by a positive relationship with medial beta power and medial and frontopolar theta power. (*b*) When participants selected the response with greater uncertainty, similar patterns emerged: frontopolar theta and mid-frontal theta correlated with relative uncertainty. This specific effect was larger in subjects that the model identified as using uncertainty to guide exploration (positive ε). (*c*) In stark contrast, when participants selected the option with the lower relative uncertainty, there were widespread negative correlations between relative uncertainty and medio/lateral/frontal beta and theta power and no difference between ε groups.

The FRN/theta signal is commonly assumed to reflect a signed negative prediction error since it is larger when stimuli indicate a loss in value or a worse-than-expected outcome (Holroyd and Coles 2002; Hajcak et al. 2006), and the amplitude scales linearly with the degree of punishment expectancy (Cavanagh et al. 2010; Chase et al. 2010; Ichikawa et al. 2010; Philiastides et al. 2010). Even when valence and expectation are independently manipulated, there appear to be separate (Pfabigan et al. 2011) or interactive (Potts et al. 2010) effects of both negative valence and expectancy in determining FRN amplitude. However, recent studies have detailed how a larger FRN has been observed to rewards when participants were expecting punishment (Oliveira et al. 2007; Baker and Holroyd 2011). Thus, it is clear that this signal can be

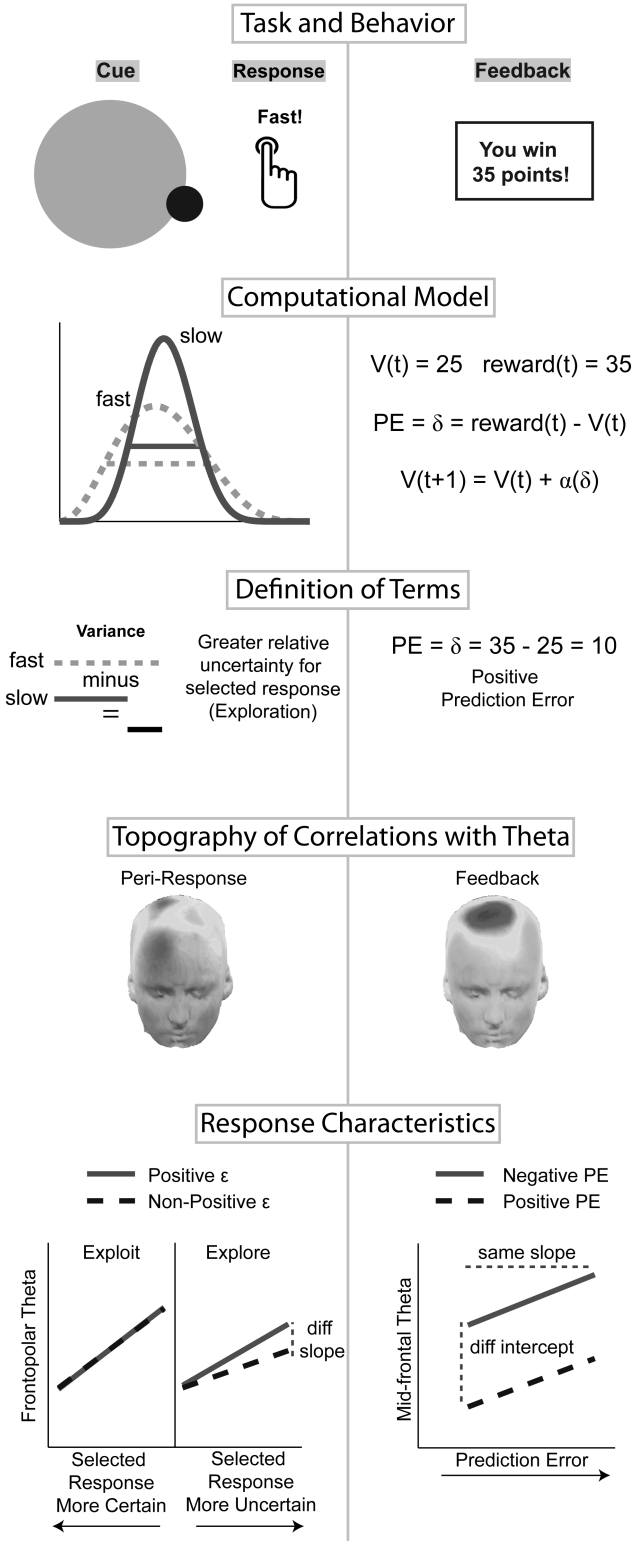

Figure 8. Summary figure. In the left column, a sample trial details that a fast response was chosen. Faster responses had higher probability distribution variance than slow responses; thus, there was more relative uncertainty for the chosen option, which was defined as an exploratory trial. Greater relative uncertainty correlated with peri-response theta power in mid-frontal and right frontopolar areas. Response characteristic slopes were taken from unthresholded TF-ROIs in Figure 7; conditions where participants selected the response with lower relative uncertainty (exploitation) were inverted from Figure 7*c* to demonstrate the continuous nature of the relationship between relative uncertainty of the selected response and theta power. This plot shows how individuals who used uncertainty to explore (positive ε) had significantly larger slopes only when exploring. The right column shows a sample feedback. Given the current estimation of action value for fast responses (25), this feedback (35) reflected a better-than-expected outcome (positive prediction error: +10). Prediction errors correlated with mid-frontal theta power. Response characteristic slopes were taken from unthresholded TF-ROIs from Figure 5; valenced prediction errors had similar slopes, yet negative prediction errors were characterized by a larger initial offset (greater power overall).

modulated by degree of unexpectedness to rewards or punishments, suggesting that it may reflect a general signal of expectation violation rather than signed reward prediction error.

This is an important distinction as Caplin and Dean (2008) have recently argued that neural responses must fulfill specific axiomatic criteria to be declared reward prediction errors. The FRN/theta signal appears to violate a core feature of these criteria: unique sensitivity to a signed signal. The current study demonstrated that feedback expectedness scaled linearly with theta power for both positive and negative prediction errors (similar slope) in the context of an asymmetrical sensitivity to negative prediction errors (higher intercept). Thus, it does not appear to reflect the same type of reward prediction error coded by dopamine neurons (Schultz 2002) and striatal activities (Rutledge et al. 2010), although it perfectly replicates patterns observed in dorsal cingulate neurons (Hayden et al. 2011). It is possible that the conjoined sensitivity of this signal to unexpectedness and negative sign reflects a dichotomized computational scheme for coding a punishment prediction error. However, it is also possible that increased theta power to negative prediction errors is related to valence or behavioral adjustment strategies. We hypothesize that future investigations will reveal separate unexpectedness updating (slope) and valence-related (intercept) features of this signal and further support the distinction of this signal apart from a reward prediction error.

This reinterpretation of the FRN/theta response as a signal of expectedness violation in the context of behavioral adaptation fits with recent theoretical and empirical work (Holroyd et al. 2008; Cavanagh et al. forthcoming). Although this finding actually diverges from our previous work describing how mid-frontal theta specifically scaled with negative prediction error (Cavanagh et al. 2010), it fits with another finding from that same paper describing how lateral theta power scaled with unsigned prediction error. In that same study, the lateral frontal theta signal preferentially predicted future behavioral adaptation in seeking reward, suggesting a specific role in reward-based behavioral optimization. The findings from the current study parallel this interpretation, where right frontal areas were involved in the suggested top-down influence over exploratory behavior.

Quantifying Exploration

Modeling exploration is not trivial because it requires predicting that participants make a response that counters their general propensity to exploit the option with highest value, and therefore, any model of exploration requires knowing "when" this will occur. Daw et al. (2006) did not find evidence that relative uncertainty about the reward statistics was associated with exploration when it was modeled as an uncertainty bonus. Rather, they found that exploration was best characterized by the standard softmax logistic choice

function in which exploration randomly occurred due to noise. Payzan-LeNestour and Bossaerts (2011) assessed the positive (bonus) or negative (penalty) influence of estimation uncertainty on exploration, discovering a better fit for negative influence—suggestive of ambiguity aversion. However, a negative influence of uncertainty may also arise simply because the majority of trials are exploitative and tend to be directed toward the more certain option. Our analysis facilitated the revelation of a positive influence of uncertainty, particularly in those trials in which participants were putatively exploring. Moreover, it is worthy to note that studies failing to report a positive effect of uncertainty on exploration have all used *n*-armed bandit tasks with dynamic reward contingencies across trials (Daw et al. 2006; Jepma and Nieuwenhuis 2010; Payzan-LeNestour and Bossaerts 2011), where participants may have responded as if only the very last trial was informative about value (as adduced from the effective very high learning rates in Daw et al. (2006) and Jepma and Nieuwenhuis (2010)). It may be difficult to estimate uncertainty-driven exploration in this context, given that participants would be similarly uncertain about all alternative options that had not been selected in the most recent trial. Here we used a task that did not have any structural uncertainty (the reward dynamics did not shift or reverse within a block, only between blocks); therefore, exploration could be used to reduce uncertainty. Indeed, we found significant positive correlations between RT swings from one trial to the next and the relative uncertainty in those trials in participants with positive ε (Fig. 3*c*).

Uncertainty-Driven Exploration and Frontal Theta

Mid-frontal and right lateral/frontopolar areas appeared to "track" the degree of relative uncertainty of the chosen option (as compared with the unchosen option) up to 500 ms prior to response commission (Fig. 7*a*). Mid-frontal response-locked ERP magnitudes have been shown to vary with uncertainty about whether the selected response was correct during action-monitoring tasks (Scheffers and Coles 2000; Pailing and Segalowitz 2004). While these aforementioned response-locked ERPs have also been shown to reflect phase-locked mid-frontal theta-band activities (Luu and Tucker 2001; Luu et al. 2004; Trujillo and Allen 2007), we are unaware of any investigations of the EEG correlates of response uncertainty during reinforcement learning. We suggest that this finding relates to uncertainty-driven exploration in this specific instance based on 3 pieces of evidence.

First, positive relationships between relative uncertainty and EEG power were only present when participants were selecting the option with greater uncertainty, which we have previously associated with exploration (Frank et al. 2009). In fact, the finding that these EEG–uncertainty relationships were positive during exploration (Fig. 7*b*) and negative during exploitation (Fig. 7*c*) suggests a direct relationship between increasing uncertainty of the chosen option and EEG power (Fig. 8). Second, these correlations were located over cortical areas previously associated with exploration (dorsomedial cortex [Pearson et al. 2009] and right frontal pole [Daw et al. 2006]) and in a frequency band previously associated with strategic control (theta) when selecting the option with greater uncertainty. In contrast, correlations with EEG power during exploitation were broad in both topography and frequency. Finally, these exploration-related effects were significantly larger in participants with a positive ε parameter (whom the model identified as using uncertainty to guide exploration) when compared with participants with a non-positive ε parameter. Convergent evidence suggests that the enhanced theta power over mid-frontal and right lateral/frontopolar areas reflects the decision to commit to a response when exploring.

Conclusions

The use of computationally derived, theoretically motivated parameters clarify the likely roles of the EEG patterns observed here. EEG activities in distinct frontal regions, primarily in the theta band, correlated with separable features of reinforcement learning: unexpectedness updating and decision uncertainty. First, we "clarify" the features of feedback-related unexpectedness updating: this signal is linearly related to unexpectedness but has an asymmetrical sensitivity to negative outcomes. Second, we provided the first evidence for a link between frontocortical activity and the strategic application of uncertainty-driven exploration.

Notes

Conflict of Interest: None declared.

References

Badre D. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends Cogn Sci. 12:193–200.

Baker TE, Holroyd CB. 2011. Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. Biol Psychol. 87:25–34.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF. 2007. Learning the value of information in an uncertain world. Nat Neurosci. 10:1214–1221.

Bush G, Vogt BA, Holmes J, Dale AM, Greve D, Jenike MA, Rosen BR. 2002. Dorsal anterior cingulate cortex: a role in reward-based decision making. Proc Natl Acad Sci U S A. 99:523–528.

Caplin A, Dean M. 2008. Axiomatic methods, dopamine and reward prediction error. Curr Opin Neurobiol. 18:197–202.

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

Cavanagh JF, Frank MJ, Klein TJ, Allen JJB. 2010. Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. NeuroImage. 49:3198–3209.

Cavanagh JF, Zambrano-Vazquez L, Allen JJB. Forthcoming. Theta lingua franca: a common mid-frontal substrate for action monitoring processes. Psychophysiology. doi: 10.1111/j.1469-8986.2011.01293.x.

Chase HW, Swainson R, Durham L, Benham L, Cools R. 2010. Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. J Cogn Neurosci. 4:936–946.

Cohen JD, McClure SM, Yu AJ. 2007. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos Trans R Soc Lond B Biol Sci. 362:933–942.

Cohen MX, Cavanagh JF. 2011. Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. Front Psychol. 2:1–12.

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

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. 2006. Cortical substrates for exploratory decisions in humans. Nature. 441:876–879.

Dayan P, Sejnowski TJ. 1996. Exploration bonuses and dual control. Mach Learn. 25:5–22.

Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK. 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci. 25:11730–11737.
Delorme A, Makeig S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 134:9–21.
Frank MJ, Doll BB, Oas-Terpstra J, Moreno F. 2009. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat Neurosci. 12:1062–1068.
Gittins JC, Jones DM. 1974. A dynamic allocation index for the sequential design of experiments. In: Gans J, editor. Progress in statisics. Amsterdam: North-Holland. p. 241–266.
Hajcak G, Moser JS, Holroyd CB, Simons RF. 2006. The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. Biol Psychol. 71:148–154.
Hanslmayr S, Pastotter B, Bauml KH, Gruber S, Wimber M, Klimesch W. 2008. The electrophysiological dynamics of interference during the stroop task. J Cogn Neurosci. 20:215–225.
Hayden BY, Heilbronner SR, Pearson JM, Platt ML. 2011. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci. 31:4178–4187.
Holroyd CB, Coles MG. 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev. 109:679–709.
Holroyd CB, Pakzad-Vaezi KL, Krigolson OE. 2008. The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. Psychophysiology. 45:688–697.
Ichikawa N, Siegle GJ, Dombrovski A, Ohira H. 2010. Subjective and model-estimated reward prediction: association with the feedback-related negativity (FRN) and reward prediction error in a reinforcement learning task. Int J Psychophysiol. 78:273–283.
Jepma M, Nieuwenhuis S. 2010. Pupil diameter predicts changes in the exploration-exploitation tradeoff: evidence for the adaptive gain theory. J Cogn Neurosci. 7:1587–1596.
Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF. 2006. Optimal decision making and the anterior cingulate cortex. Nat Neurosci. 9:940–947.
Koechlin E, Summerfield C. 2007. An information theoretical approach to prefrontal executive function. Trends Cogn Sci. 11:229–235.
Luu P, Tucker DM. 2001. Regulating action: alternating activation of midline frontal and motor cortical networks. Clin Neurophysiol. 112:1295–1306.
Luu P, Tucker DM, Derryberry D, Reed M, Poulsen C. 2003. Electrophysiological responses to errors and feedback in the process of action regulation. Psychol Sci. 14:47–53.
Luu P, Tucker DM, Makeig S. 2004. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. Clin Neurophysiol. 115:1821–1835.
Miltner WHR, Braun CH, Coles MGH. 1997. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. J Cogn Neurosci. 9:788–798.
Moustafa AA, Cohen MX, Sherman SJ, Frank MJ. 2008. A role for dopamine in temporal decision making and reward maximization in parkinsonism. J Neurosci. 28:12294–12304.
Oliveira FT, McDonald JJ, Goodman D. 2007. Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. J Cogn Neurosci. 19:1994–2004.
Pailing PE, Segalowitz SJ. 2004. The effects of uncertainty in error monitoring on associated ERPs. Brain Cogn. 56:215–233.
Payzan-LeNestour E, Bossaerts P. 2011. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS Comput Biol. 7:e1001048.
Pearson JM, Hayden BY, Raghavachari S, Platt ML. 2009. Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. Curr Biol. 19:1532–1537.
Pfabigan DM, Alexopoulos J, Bauer H, Sailer U. 2011. Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. Psychophysiology. 48:656–664.
Philiastides MG, Biele G, Vavatzanidis N, Kazzer P, Heekeren HR. 2010. Temporal dynamics of prediction error processing during reward-based decision making. Neuroimage. 53:221–232.
Potts GF, Martin LE, Kamp SM, Donchin E. 2010. Neural response to action and reward prediction errors: comparing the error-related negativity to behavioral errors and the feedback-related negativity to reward prediction violations. Psychophysiology. 48:218–228.
Rutledge RB, Dean M, Caplin A, Glimcher PW. 2010. Testing the reward prediction error hypothesis with an axiomatic model. J Neurosci. 30:13525–13536.
Scheffers MK, Coles MG. 2000. Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. J Exp Psychol Hum Percept Perform. 26:141–151.
Schultz W. 2002. Getting formal with dopamine and reward. Neuron. 36:241–263.
Shima K, Tanji J. 1998. Role for cingulate motor area cells in voluntary movement selection based on reward. Science. 282:1335–1338.
Strauss GP, Frank MJ, Waltz JA, Kasanova Z, Herbener ES, Gold JM. 2011. Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. Biol Psychiatry. 69:424–431.
Trujillo LT, Allen JJ. 2007. Theta EEG dynamics of the error-related negativity. Clin Neurophysiol. 118:645–668.
Wang C, Ulbert I, Schomer DL, Marinkovic K, Halgren E. 2005. Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. J Neurosci. 25:604–613.
Yoshida W, Ishii S. 2006. Resolution of uncertainty in prefrontal cortex. Neuron. 50:781–789.