Frontal Theta Oscillatory Activity Is a Common Mechanism for the Computation of Unexpected Outcomes and Learning Rate

Ernest Mas-Herrero[1] and Josep Marco-Pallarés[1,2]

Abstract

■ In decision-making processes, the relevance of the information yielded by outcomes varies across time and situations. It increases when previous predictions are not accurate and in contexts with high environmental uncertainty. Previous fMRI studies have shown an important role of medial pFC in coding both reward prediction errors and the impact of this information to guide future decisions. However, it is unclear whether these two processes are dissociated in time or occur simultaneously, suggesting that a common mechanism is engaged. In the present work, we studied the modulation of two electrophysiological responses associated to outcome processing—the feedback-related negativity ERP and frontocentral theta oscillatory activity—with the reward prediction error and the learning rate. Twenty-six participants performed two learning tasks differing in the degree of predictability of the outcomes: a reversal learning task and a probabilistic learning task with multiple blocks of novel cue–outcome associations. We implemented a reinforcement learning model to obtain the single-trial reward prediction error and the learning rate for each participant and task. Our results indicated that midfrontal theta activity and feedback-related negativity increased linearly with the unsigned prediction error. In addition, variations of frontal theta oscillatory activity predicted the learning rate across tasks and participants. These results support the existence of a common brain mechanism for the computation of unsigned prediction error and learning rate. ■

INTRODUCTION

In our daily life, we face decisions and evaluate their consequences to obtain information about how to act in similar situations in the future. Determining the value of a decision in different contexts is a complex issue, which is influenced by new evidences that are continuously collected and by the learning history. The relevance of both history and new information in guiding decision-making is influenced by the characteristics of the environment. In uncertain environments, new pieces of information have greater importance in the adaptation of behavior. In contrast, in stable environments, past experience is more relevant than recently acquired information. Therefore, we constantly evaluate how accurate our predictions are and how relevant incoming information is according to the present context to update future estimates.

Several studies have revealed a crucial role of the medial pFC (mPFC) in both action monitoring and updating of action values (Rushworth, Walton, Kennerley, & Bannerman, 2004). Specifically, it has been proposed that the mPFC monitors behavior on the bases of reward prediction errors (RPEs; discrepancies between expected and real outcomes), a process described by the principles of reinforcement learning (RL) theory (Jocham, Neumann, Klein, Danielmeier, & Ullsperger, 2009; Sutton & Barto, 1998). In addition, fMRI studies have suggested that mPFC also encodes the rate at which new information replaces outdated evidence (Jocham et al., 2009; Behrens, Woolrich, Walton, & Rushworth, 2007; Walton, Croxson, Behrens, Kennerley, & Rushworth, 2007; Yoshida & Ishii, 2006). These studies have shown that activity in the mPFC, specifically in the ACC, increases in situations in which newly acquired information is highly relevant to optimize goal-directed behavior, such in uncertain environments. This information is indexed in RL models by the learning rate parameter (α). This parameter is greater in uncertain or volatile environments than in stable contexts. In addition, variations in ACC activity during outcome monitoring predict the α values across participants, reflecting the relationship between mPFC and the updating of new information (Jocham et al., 2009; Behrens et al., 2007).

However, because of the low temporal resolution of the fMRI technique, it is still an open question whether these two processes, monitoring of behavior and updating of action values, are dissociated or not in the mPFC. The goal of this study is to determine whether computation of prediction error and determination of the learning rate are two independent neural processes or engage a common mechanism. To reach this goal, we will take

[1]L'Hospitalet de Llobregat, Barcelona, Spain, [2]University of Barcelona

Journal of Cognitive Neuroscience 26:3, pp. 447–458
doi:10.1162/jocn_a_00516

advantage of the high temporal resolution of EEG. Previous studies have described two electrophysiological responses during outcome processing, the feedback-related negativity (FRN) ERP (Gehring & Willoughby, 2002) and the mediofrontal theta oscillatory activity (Marco-Pallares et al., 2008; Cohen, Elger, & Ranganath, 2007). Previous studies using intracranial recording and source modeling have suggested that these two signals are generated in the mPFC (Cohen, Ridderinkhof, Haupt, Elger, & Fell, 2008; Luu, Tucker, & Makeig, 2004; Luu, Tucker, Derryberry, Reed, & Poulsen, 2003). Both signals peak around 250–300 msec after outcome delivery and are modulated by the degree of discrepancy between expected and real outcome (Ferdinand, Mecklinger, Kray, & Gehring, 2012; Cavanagh, Figueroa, Cohen, & Frank, 2011; Chase, Swainson, Durham, Benham, & Cools, 2011; Philiastides, Biele, Vavatzanidis, Kazzer, & Heekeren, 2010; Oliveira, McDonald, & Goodman, 2007; Holroyd & Coles, 2002). However, at present there are no studies addressing the modulation of these components by the learning rate.

In the present work, we used brain ERPs and time frequency (TF) decomposition of EEG data to study the neuropsychological markers of both the RPE and the learning rate. To reach this goal, the participants performed two probabilistic learning (PL) tasks: a reversal learning (RVL) task in which they had to adapt their behavior to unexpected changes in the environment and a PL task, which consisted of multiple blocks of novel cue–outcome associations without unexpected reversal rules. In both tasks, electrophysiological responses were analyzed based on the characteristics of a computational RL model. We hypothesized that FRN and theta oscillatory activity would be modulated by RPE in both tasks. Additionally, if these two signals are also the neural signatures of the learning rate, they should vary across participants and tasks (e.g., increasing in more uncertain environments such as the RVL task compared with the PL task).

METHODS

Participants

Twenty-six students ($M = 21.7$ years, $SD = 2.7$ years, 13 men) participated in the experiment. All participants were paid €10 per hour and a monetary bonus depending on their performance. All participants gave written informed consent, and all procedures were approved by the local ethics committee.

Experimental Procedure

Each participant performed two experimental tasks; the presentation order was counterbalanced across participants. The first was a RVL task adapted from Cools, Clark, Owen, and Robbins (2002), which consisted of 637 trials divided into 49 blocks (10–16 trials each). In each trial, two geometric figures were presented on either side of a central fixation point. The participants were instructed to select one of the figures. After a delay of 1000 msec, one of two possible types of feedback was displayed: a green tick (reward, +€0.04) or a red cross (punishment, −€0.04; Figure 1). On each block, one figure was rewarded in 75% of the trials, whereas the other was rewarded in 25% of the trials. However, at the beginning of each block, the rule was reversed. During the first five trials following the contingency reversal, a

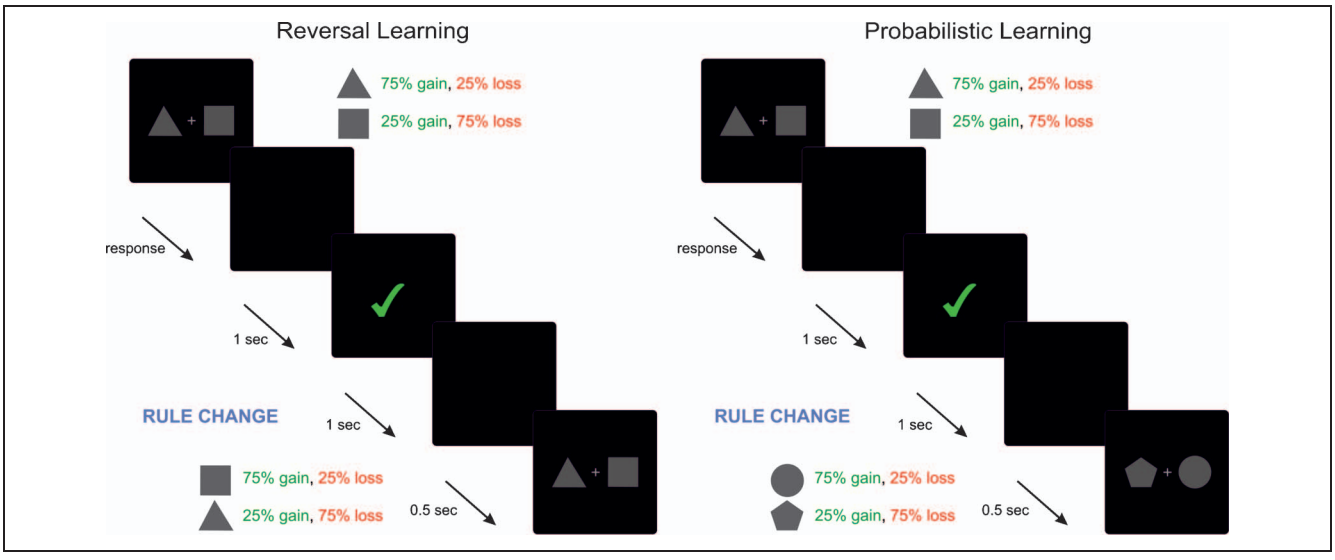

Figure 1. RVL (left) and PL (right) tasks used in the study. Both tasks consisted of 637 trials divided into 49 blocks varying from 10 to 16 trials each. On each trial, participants had to select between two geometric figures. Using trial-and-error feedback, participants had to discover the most advantageous figure. After each block, the rule changed. In the RVL, rule changes were not informed, whereas in the PL, rule changes were indicated by the presentation of two new figures.

selection of the previously correct stimulus would result in punishment.

The second task was a PL task, which also consisted of 637 trials divided into 49 blocks of 10–16 trials. As in the RVL, participants had to choose between two geometric figures that were rewarded differently (75% vs. 25% rewarded), resulting in two possible feedbacks: a green tick (reward, +€0.04) or a red cross (punishment, −€0.04). However, in this task, there were no uninformed reversal contingencies. After each block (10–16 trials), two new figures were presented. Therefore, in each block, the participants had to discover the rule that would remain constant for the remainder of the block. During the first five trials of each block, a selection of the incorrect stimulus would lead to punishment. Blocks were not followed by breaks or pauses; here we refer to block as periods in which cue–outcome associations remain stable. The duration of the stimulus presentation was the same as in RVL (Figure 1). Both tasks were preceded by a short training session.

In both tasks, if the participants did not respond in the requested time (1000 msec), a question mark appeared on the screen after the stimuli. These trials were discarded from further analysis. Self-paced rest periods were given after 35–40 trials. During these pauses, the participants were told how much money they had earned up to that point. The participants were encouraged to earn as much money as possible in both tasks. The participants were explicitly informed that one task involved uninformed reversals (RVL) and the other did not (PL).

Electrophysiological Recording

EEG was recorded from the scalp (0.01 Hz high-pass filter with a notch filter at 50 Hz; 250 Hz sampling rate) using a BrainAmp amplifier with tin electrodes mounted in an electrocap (Electro-Cap International) located at 29 standard positions (Fp1/2, Fz, FCz, F7/8, F3/4, Fc1/2 Fc5/6, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T5/6, PO1/2, Oz) and the left and right mastoids. An electrode placed at the lateral outer canthus of the right eye served as an online reference. EEG was rereferenced offline to the linked mastoids. Vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. Electrode impedances were kept below 5 kΩ. Trials with absolute mean amplitudes higher than 100 μV were automatically rejected offline. Six participants were excluded from the study because they had trial rejection rates higher than 20%.

EEG Analysis

The FRN was studied by epoching EEG data from 100 msec time-locked before the outcome (baseline) to 600 msec after the outcome onset. Following previous studies (Gehring & Willoughby, 2002), FRN was analyzed by averaging the amplitude in a time window located 40 msec around the peak, which was located between 240 and 300 msec for each experimental condition at FCz. However, this mean amplitude is affected by the concomitant P300, which we hypothesized might respond differently to experimental conditions. To minimize this effect, ERP epochs were first high-pass-filtered at 3 Hz to remove slow-frequency noise such as P300 (Wu & Zhou, 2009).

Time–frequency analysis was performed per trial in 4-sec epochs (2 sec before feedback through 2 sec after) using seven-cycle complex Morlet wavelets. Considering previous studies (Marco-Pallares et al., 2008; Cohen et al., 2007), we specifically focused on theta (5–7 Hz), which has been implicated in both reward and punishment processing. To analyze trial-by-trial modulations, we computed changes in time-varying energy (square of the convolution between wavelet and signal) in the studied frequencies with respect to baseline for each trial. To compare different conditions, trials associated with a specific condition were averaged for each participant before performing a grand average. Following previous studies (Cavanagh, Zambrano-Vazquez, & Allen, 2012; Luu et al., 2004), the mean increase/decrease in power for each condition was computed at FCz.

RL Model

A Q-learning model used by Watkins and Dayan (1992) was implemented in both tasks. The model used RPE to update the weights associated with each stimulus and probabilistically chose the stimulus with the higher weight. The weight was then updated using the following algorithm:

$$W(t+1) = W(t) + \alpha \cdot \delta$$

where α is the learning rate and δ represents the prediction errors, calculated as the difference between the outcome and the expectancy or weight of the selected figure. Next, softmax action selection was used to compute the probability of choosing one of the two options:

$$P_A(t) = \frac{e^{\gamma \cdot W_A(t)}}{e^{\gamma \cdot W_A(t)} + e^{\gamma \cdot W_B(t)}}$$

where γ is an exploitation parameter (the inverse of the temperature parameter).

The model was run 10 times using random initial values for each participant by maximizing the log-likelihood estimate (LLE). We used the fminsearch function of Matlab R2008, which uses a Nelder–Mead simplex method (Cohen & Ranganath, 2007). The parameters α and γ with the best LLE were selected. The model was run across the entire task in the RVL task. On the other hand, in the PL task, the model was run for each block, that is, for each new cue–outcome association. In the PL, those blocks

in which the difference between the LLE derived from the model and the LLE of a chance performance model was less than 3—suggesting a poor fit—were discarded for the analysis (M = 17%, SD = 0.7%; Kass & Raftery, 1995). Once α and γ were individually calculated, values representing the prediction error could be determined on a trial-by-trial basis. Finally, we also computed, for each participant, the probability of choice predicted by the model on each trial considering the parameter computed (α and γ), the participant's responses, and the feedback delivered. To show the consistency of model's prediction in both tasks, we plotted the average of both real participants' choice and the probability of choice computed by the model across trials. In addition, we also computed the probability of choice given the mean of α and γ of all participants and using simulated data. If the model fits well, participants' choice should match the probability of choice predicted by the model using both participants' behavior and simulated data.

Statistical Analysis

To study which components of ERP and TF during feedback evaluation were associated with the prediction errors extracted from the model, negative and positive trials were independently sorted into three bins according to the size of the absolute RPE: those with high (HPE), medium (MPE), and low (LPE) prediction error (with each group defined by the 33rd, 66th, and 100th percentile of the range).

In both tasks, differences among conditions in both ERP and TF data were determined by repeated-measures ANOVA with two within-participant factors: Valence (positive and negative) and Absolute RPE (high, medium, and low).

In the RVL task, in addition, regression analysis was performed using absolute RPE as a predictor of FRN amplitude and oscillatory activity. We then determined whether the value of the slope was different overall from 0 for the group for RPE measure using a one-sample t test. A significant difference from 0 would suggest a relationship between the size of the prediction error and the size of the FRN amplitude or TF activity. Separate analyses for positive and negative prediction errors were also performed.

In PL, the amplitude of the FRN and theta activity within trials may be modulated not only by RPE but also by the difference of learning rate among blocks. To test this hypothesis, we performed a multiple regression analysis with two independent measures: absolute RPE and the learning rate associated to each block. Again, separate analysis for positive and negative feedback were performed.

We used Spearman correlation to study the relationship between participant's learning rate and both midfrontal theta activity and FRN amplitude during RVL tasks. Finally, we studied whether differences in learning rate between tasks and across participants may predict differences in FRN amplitude and oscillatory activity. For that reason, we performed Spearman correlations of overall FRN amplitude and theta activity with the difference of the learning rates obtained in the RVL and the PL. To obtain a unique learning rate for each participant in the PL to compare it with the learning rate obtained in the RVL task, we average the learning rates obtained across blocks for each individual. We performed separate analyses for positive and negative feedback. Participants with theta activity and FRN amplitude greater than 2.5 SD in any of the conditions were not included in the correlation analysis of each specific condition.

For all statistical effects involving two or more degrees of freedom in the numerator, the Greenhouse–Geisser epsilon was used as needed to correct for possible violations of the sphericity assumption. The p values following correction are reported.

RESULTS

RVL Task

The participants selected the most rewarded figure in 77% (SD = 4.5%) of the trials with a mean RT of 446.01 msec (SD = 63.79 msec) and performed a switch after 2.3 (SD = 0.5) consecutive negative outcomes. Previous studies have shown similar error perseverance in this task (Chase et al., 2011). The participants earned €7.35 (SD = €1.2) on average.

The RL model was fitted to participants' behavioral performance (pseudo-R^2 = .48, SD = .12). Participants had a mean learning rate of 0.62 (SD = 0.2) and a mean exploitation parameter of 0.27 (SD = 0.04). Figure 2A shows an example of the behavior of one participant and the predictions generated by the model with the parameters estimated for this individual by the RL model (α and γ) and participant's data. The model successfully predicts most of the responses generated by the participant. Additionally, Figure 2B shows percentage of participant's choice as well as the predictions generated by the model based on participants' behavior and simulated data. Although model prediction matched most of participants' responses, model predictions and participants' behavior did not fully match between Trials 3 and 6. These differences could be because of other different learning systems operating in parallel-like model-based learning (Gläscher, Daw, Dayan, & O'Doherty, 2010).

Mean amplitudes of the FRN for trials with high, medium, and low absolute RPE were extracted in both positive and negative feedback and analyzed by repeated-measures ANOVA. Figure 3 shows that feedback induced a negative waveform around 260–300 msec (FRN), which was more pronounced in negative than in positive feedbacks (valence effect, $F(1, 19) = 12.0$, $p < .01$). Topographical maps (see Figure 3) revealed that this effect was maximal at FCz. Additionally, we found a significant linear effect of

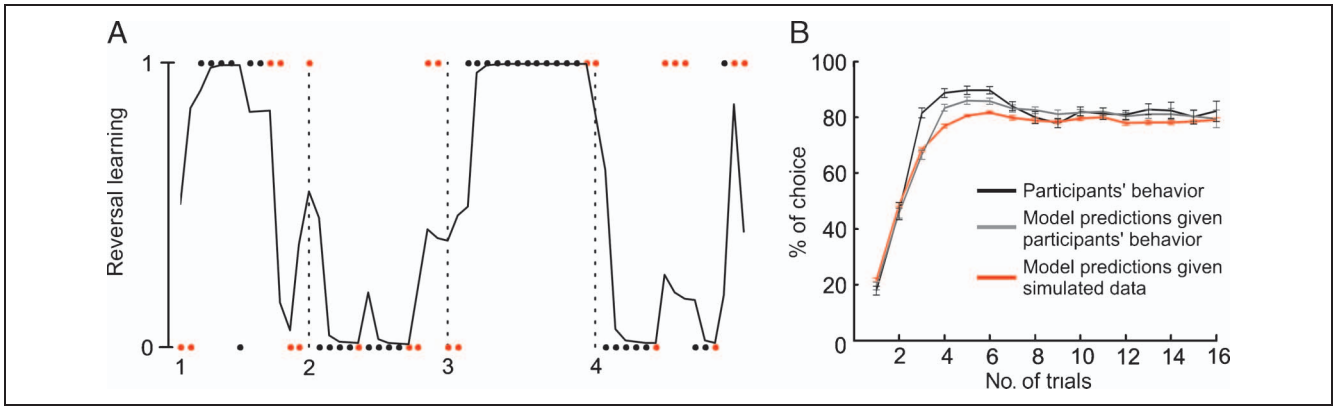

Figure 2. (A) Example of one participant's behavior in the RVL task. The participant's learning rate in RVL task was 0.66. Black dots indicate that the participant received a positive feedback, whereas red dots indicate a negative feedback. Dots in the upper part of the panel indicate that the participant selected Stimulus A, whereas dots in the lower part of the graph indicate that participant selected the Stimulus B. Dashed lines indicate the beginning of a new block, which in RVL indicates a rule reversal. Solid lines indicate the probability of selecting Stimulus A according to the RL model predictions. (B) Learning curves during RVL. The trial number is represented on the *x* axis. The values on the *y* axis are the percentage of trials in which the most rewarding stimulus was selected. Black lines show participants' behavior; gray lines indicate the prediction of the model given participant's data; red lines indicate the prediction of the model with simulated data. Error bars represent the *SEM*.

RPE, $F(1, 19) = 16.9$, $p = .001$, which was not affected by valence (RPE × Valence, $F(1, 19) = 2.4$, $p = .13$). Therefore, feedback associated with high absolute RPE elicited a more negative deflection of the FRN than trials associated with low absolute RPE in both positive and negative feedback. These results suggest that FRN amplitude does not only reflect negative RPE but also increases linearly with outcomes' expectancy deviation independently of the valence. Therefore, the FRN amplitude is also modulated by an unsigned RPE, that is, when something is different (rather than worse or better) than expected.

To test this relationship between absolute RPE and FRN amplitude, we performed a regression analysis, with all the trials for each participant, using FRN amplitude as

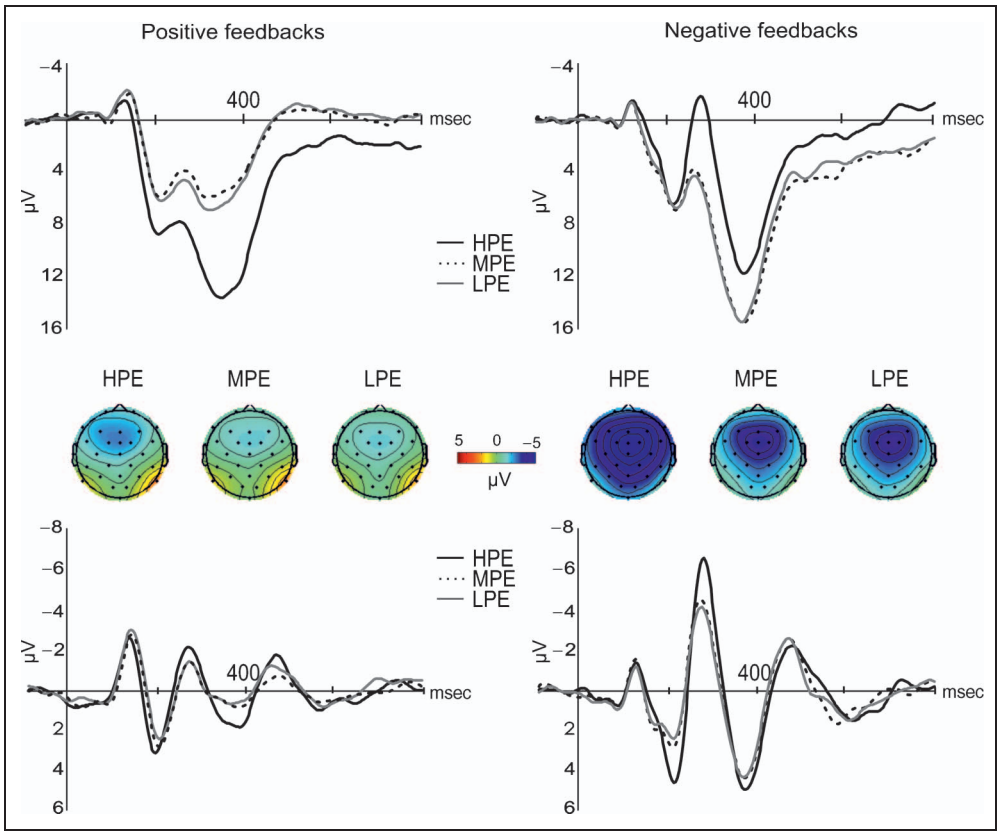

Figure 3. ERPs and topographical mapping for each condition in the RVL tasks. (Top) The ERPs without high-pass filter and (bottom) the ERPs with 3-Hz high-pass filter. (Middle) Topographical maps for the six conditions studied (260–300 msec after feedback). Note that the maximum activity for the FRN was located at FCz.

dependent variable and absolute RPE as independent measure. As suggested in the previous analysis, there was a negative relationship between these two measures in all participants. Higher FRN amplitude (more negative deflection) was associated to increase in the size of the absolute RPE. The mean slope for the group was significantly different from zero, $t(19) = -3.9$, $p = .001$. We also repeated the same analysis separately for positive and negative feedbacks. As expected, there was a negative relationship, and the mean slope was also significantly different from 0 in positive, $t(19) = -3.7$, $p < .01$, and negative feedbacks, $t(19) = -3.3$, $p = .001$.

The time–frequency analysis of the six conditions (high, medium, and low RPE for both positive and negative feedbacks) revealed a clear enhancement of theta activity (5–7 Hz) between 100 and 600 msec after feedback onset (Figure 4). The maximum of activity was found between 280 and 400 msec, and this was the time window chosen for further analysis. This enhancement of theta power increase was more pronounced in negative trials, $F(1, 19) = 17.8$, $p < .001$, and increased linearly with RPE, $F(1, 19) = 7.6$, $p < .05$. However, the two main effects did not interact ($F < 1$). These results suggest that, as FRN, theta activity is also modulated according to unsigned RPE. We repeated the previous regression analysis using theta activity, instead of FRN amplitude, as dependent measure. There was a positive relationship between both measures in all but one of the participants and the mean slope significantly differed from 0, $t(19) = 3.8$, $p = .001$. The same results were obtained when positive feedbacks were analyzed separately, $t(19) = 3.2$, $p < .01$. A trend toward a significant effect in negative feedbacks was also found, $t(19) = 1.9$, $p = .08$.

Finally, we computed the overall theta and FRN amplitude during the entire task and correlated it with the participants' learning rates. The analysis revealed a significant positive correlation between α and theta power, $\rho(20) = .59$, $p < .01$, but not with the FRN, $\rho(20) = -.04$, $p = .88$. We performed the same analysis separately for positive and negative feedback to study whether the relationship between midfrontal theta activity and learning rate was independent from valence or, in contrast, was only present in one type of feedback (Figure 5A, B). In both cases, individual differences in learning rate predicted individual differences in theta activity (positive, $\rho(20) = .57$, $p < .01$; negative, $\rho(18) = .58$, $p = .01$). No significant correlation was found with the FRN in any case (positive, $\rho(20) = -.08$, $p = .74$; negative, $\rho(18) = -.02$, $p = .94$).

In summary, FRN and theta oscillatory activity were modulated according to an unsigned RPE, and additionally, individual differences in frontal theta oscillatory activity predicted the learning rate across participants.

PL Task

The participants selected the most rewarded figure 89.3% of the time ($SD = 3.4\%$), with a mean RT of 448.47 msec

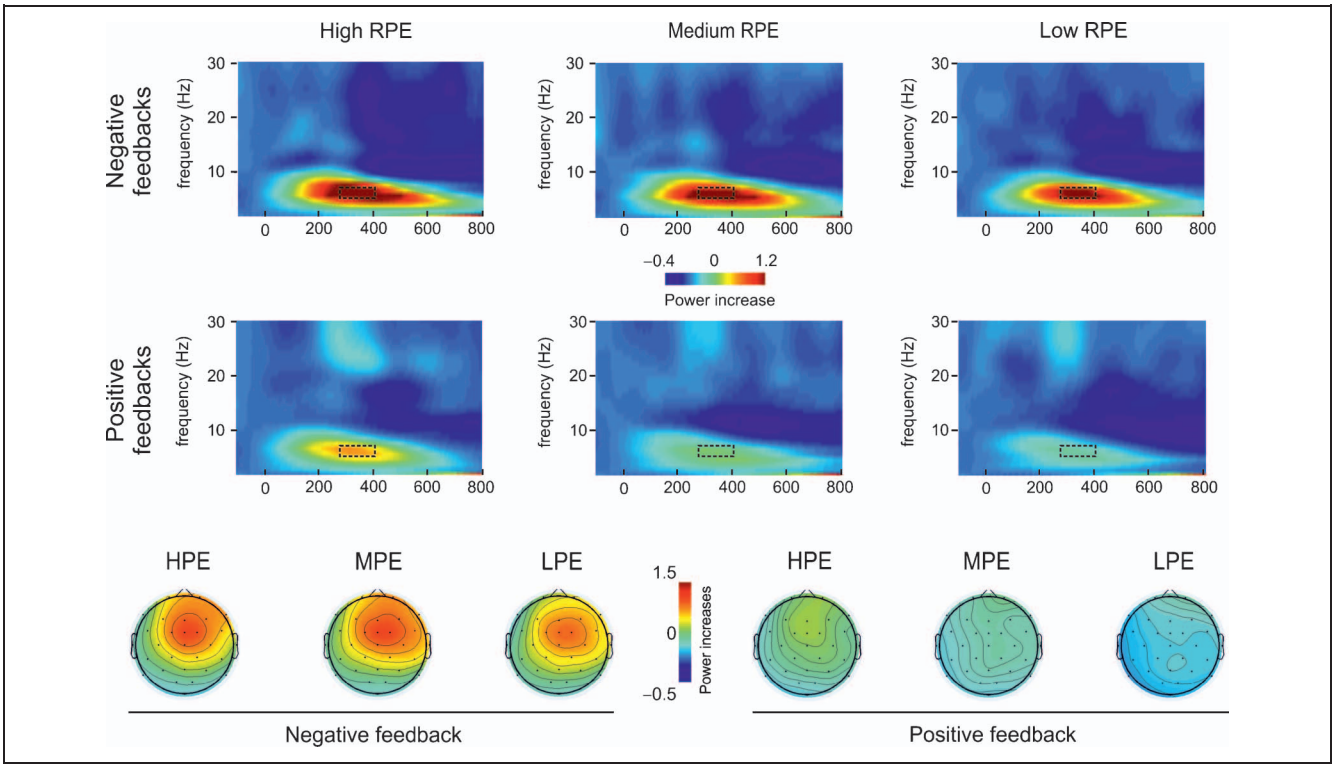

Figure 4. Changes in power at FCz with respect to baseline (100 msec before feedback onset) for negative (top) and positive feedbacks (bottom) according to the prediction error (HPE left, MPE medium, LPE right) in the RVL task.

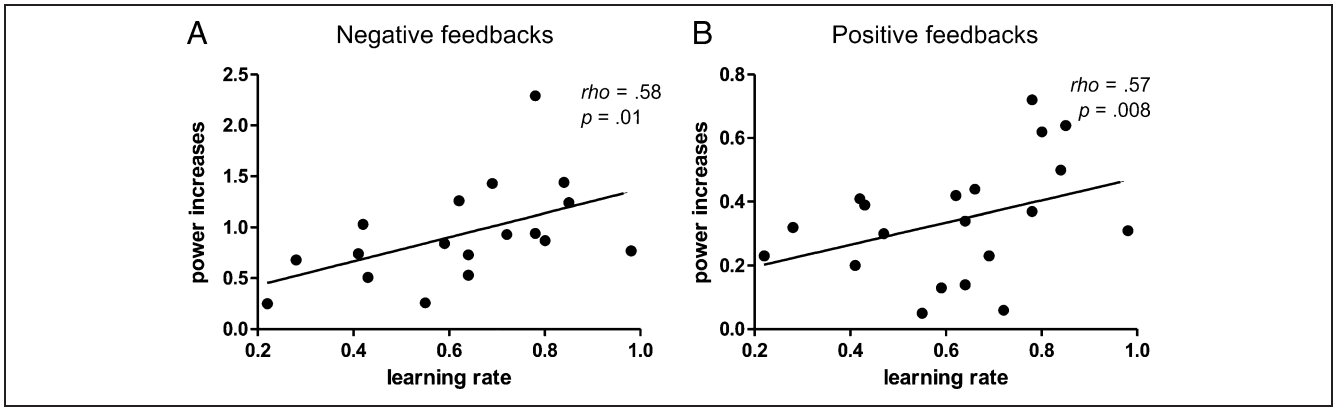

Figure 5. Scatter plot of theta power increase of negative (A) and positive feedbacks (B) and the participants' learning rate in the RVL task.

(*SD* = 56.38 msec). Behaviorally, all of the participants quickly adapted their decision-making to maximize rewards. The participants reached the most rewarded figure after 1.3 (*SD* = 0.2) trials of negative feedback at the beginning of the block. At the end of the task, the participants accumulated €11.03 (*SD* = €1.4) on average.

The RL model was fitted to participants' behavioral performance for each block (pseudo-R^2 = .77, *SD* = .06). Participants had a mean learning rate of 0.35 (*SD* = 0.04) and a mean exploitation parameter of 22.72 (*SD* = 4.06). Figure 6A shows an example of the behavior of one participant and the predictions generated by the model given participant's data. We have selected four blocks with different learning rates (Block 1 = 0.27, Block 2 = 0.98, Block 3 = 0.35, Block 4 = 0.07). In Blocks 1, 2, and 4, the participant received a punishment after selecting a stimulus that was previously rewarded for three to four times. However, participants' behaviors varied across blocks. In the second block, the participant immediately selected the second stimulus, whereas in Blocks 1 and 4, the participant perseverated after three or four punishments more, respectively. That is, in the second block, one simple punishment was enough to decrease the value of the selected compared with the unselected stimulus, whereas in the fourth block, four negatives feedbacks were required to reach such threshold. These differences in behavior had their parallel in the learning rate computed for each block, with a high learning rate in the second block (0.98) and a low learning rate in the fourth (0.07). Similarly, Figure 6B also shows that model predictions matched most of the choices performed by the participants.

In general, participants presented a smaller learning rate, $t(19) = 14.81$ $p < .001$, but a higher exploitation parameter, $t(19) = 24.87$ $p < .001$, in PL than in RVL. These differences were expected, as that in the PL task, once the correct figure has been found, participants hardly change their selection (Figure 6B). The learning rate has been suggested to be modulated according to

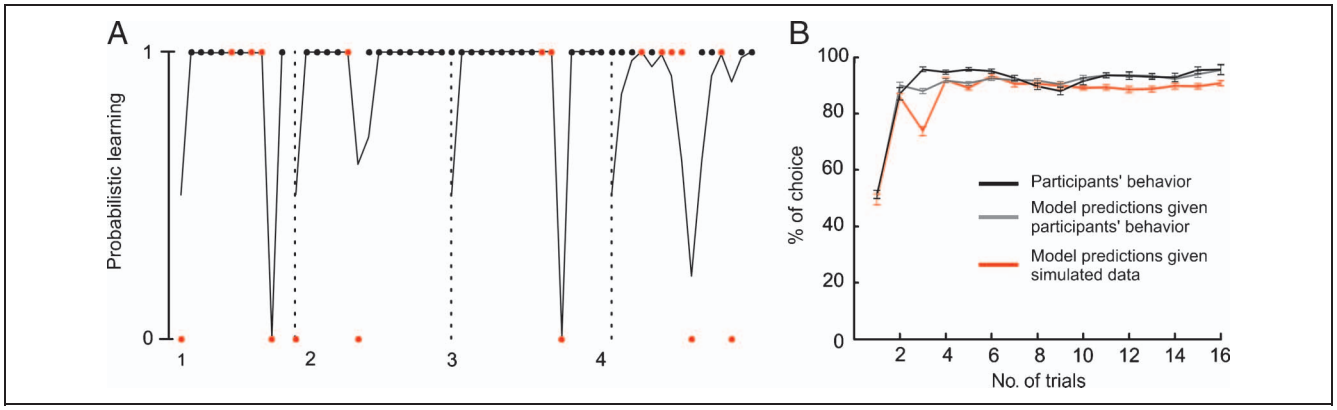

Figure 6. (A) Example of one participant's behavior in the PL task. Four blocks with different learning rates were selected (Block 1 = 0.27, Block 2 = 0.98, Block 3 = 0.35, Block 4 = 0.07). Black dots indicate that the participant received a positive feedback, whereas red dots indicate a negative feedback. Dots in the upper part of the panel indicate that the participant selected Stimulus A, whereas dots in the lower part of the graph indicate that the participant selected the stimulus B. Dashed lines indicate the beginning of a new block, which indicates the presentation of new cues. Solid lines indicate the probability of selecting the Stimulus A according to the RL model predictions. (B) Learning curves during PL. The trial number is represented on the *x* axis. The values on the *y* axis are the percentage of trials in which the most rewarding stimulus was selected. Black lines show participants' behavior; gray lines indicate the prediction of the model given participant's data; red lines indicate the prediction of the model with simulated data. Error bars represent the *SEM*.

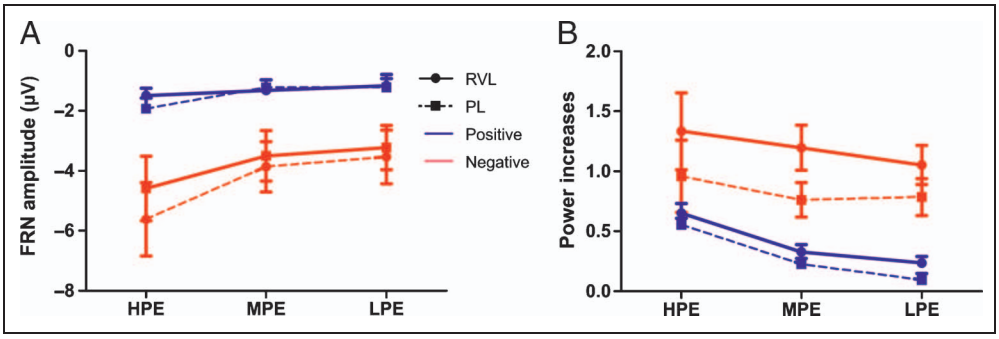

Figure 7. (A) FRN power and (B) Theta in trials presenting high, medium, and low prediction error for positive and negative feedback according to the RL model for both the RVL and PL tasks.

environmental uncertainty (Behrens et al., 2007). In that sense, the RVL task includes an extra source of uncertainty compared with PL: the rule uncertainty (rule changes were unpredictable). In contrast, in the PL task, no unpredictable changes occur within blocks. In uncertain situations, new information becomes more relevant (high learning rate) than in stable environments (low learning rate). This difference in uncertainty also affects participants' perseverance, which explains differences in the exploitation parameter.

Mean amplitudes of the FRN for trials with high, medium, and low absolute RPE within each block were extracted for both positive and negative feedback and analyzed by repeated-measures ANOVA. Similar to the results obtained in the RVL tasks, FRN amplitude was more pronounced in negative than in positive feedback (Valence effect, $F(1, 19) = 10.3, p < .005$) and scale linearly with RPE, $F(1, 19) = 3.8, p < .05$, independent of feedback valence (RPE × Valence, $F(1, 19) = 2.02, p = .15$; Figure 7A).

To study as well how the learning rate may modulate FRN amplitude, we performed a regression analysis with the FRN amplitude as dependent variable and absolute RPE and learning rate as independent variables. As it was expected from the previous analysis, the mean slope of absolute RPE was significantly different from 0, $t(19) = -2.2, p < .05$; $t(19) = -2.9, p < .01$, in both positive and negative feedbacks, respectively. However the mean slope of the learning rate ($t < 1$; $t(19) = -1.2$, $p = .23$) in both positive and negative feedback was not significantly different from 0. Thus, FRN is modulated by unsigned RPE but is not affected by the learning rate.

The time–frequency analysis of the six conditions (high, medium, and low absolute RPE for both positive and negative feedbacks) revealed a clear enhancement of theta

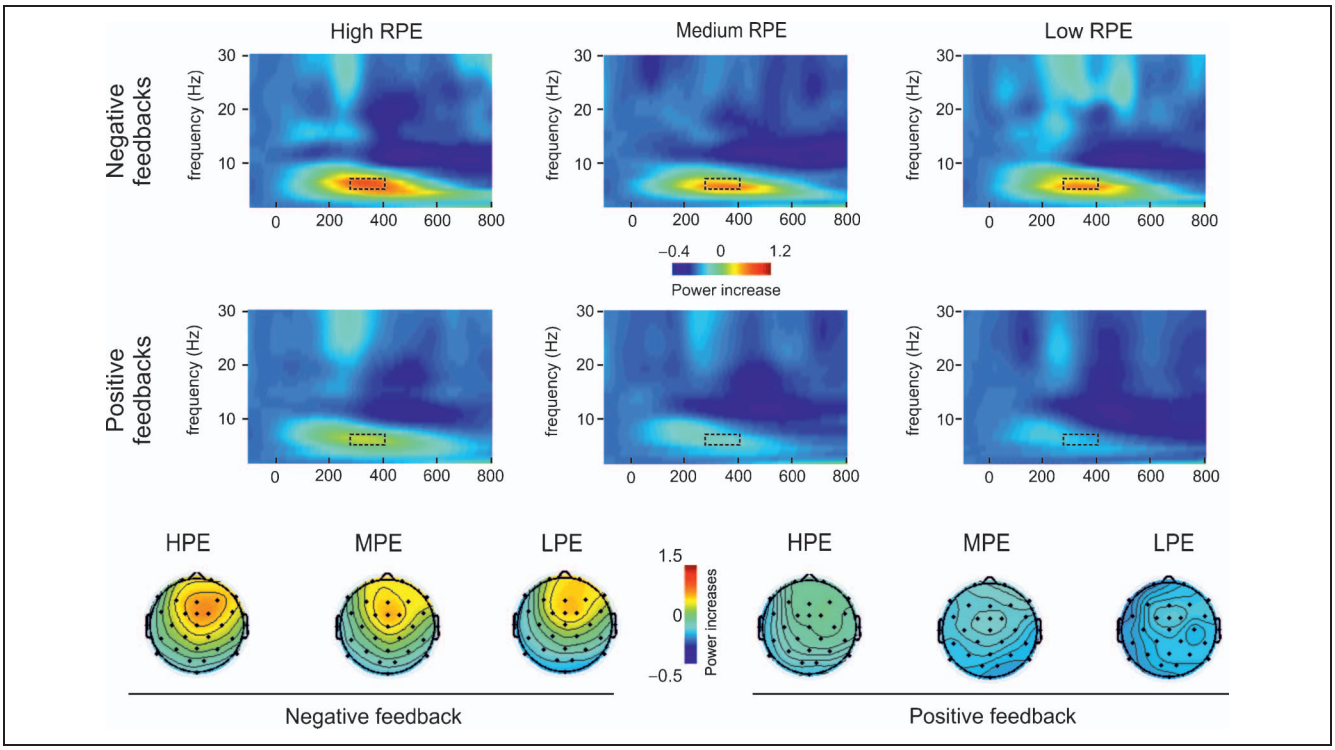

Figure 8. Changes in power at FCz with respect to baseline (100 msec before feedback onset) for negative (top) and positive feedbacks (bottom) according to the prediction error (HPE left, MPE medium, LPE right) in the PL task. Scalp maps in the dashed box are also represented.

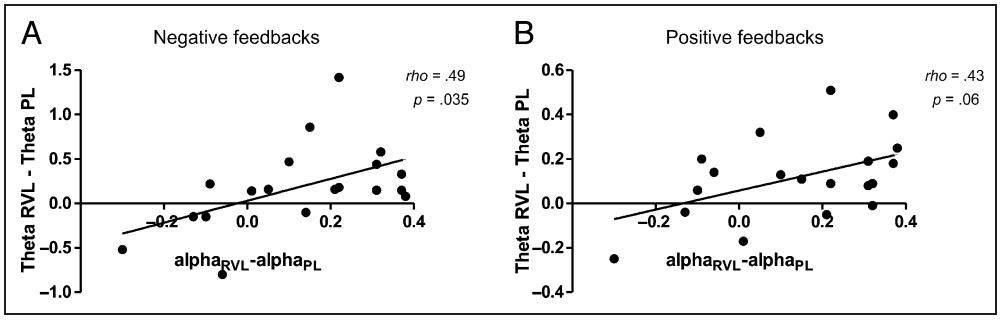

Figure 9. Scatter plot of differences in theta activity between both tasks for both negative (A) and positive (B) feedback and differences in their learning rate. The solid black line represents the slope of the linear fit.

activity (5–7 Hz) for negative trials, $F(1, 19) = 8.8, p < .01$. Additionally, it increased linearly with RPE, $F(1, 19) = 10.4$, $p < .001$, independent of the feedback valence, $F(1, 19) = 1.6, p = .22$ (Figures 7B and 8).

However, considering the results obtained in RVL, theta oscillatory activity should be modulated by both the absolute RPE values and the different learning rates obtained in each block. To test this hypothesis, we performed a regression analysis with theta oscillatory activity as dependent variable and absolute RPE and block's learning rate as independent measures. As previously reported, there was a positive relationship between absolute RPE and theta activity in all but two participants. Additionally, the learning rate was also positively related with theta activity in 18 participants. Mean slopes for all participants differed from 0 (RPE, $t(19) = 4.2, p < .001$; Alpha, $t(19) = 3.5, p < .01$). This effect was also significantly different from 0 when positive feedbacks were separately analyzed (RPE: $t(19) = 3.3, p < .01$; Alpha: $t(19) = 2.5, p < .05$), whereas for negative feedbacks, results were marginally significant (RPE: $t(19) = 1.6, p = .12$; Alpha: $t(19) = 1.6, p = .12$).

As stated before, the participants showed higher learning rates during RVL than in PL. Frontal theta oscillatory activity was also higher in RVL than in PL task, $t(19) = 2.7$, $p < .05$ (Figure 7B), suggesting a relationship between this component and the learning rate. However, these differences were significant for positive feedback, $t(19) = 4.5, p < .001$, but only marginal for negative, $t(19) = 1.8$, $p = .09$, feedback. Finally, differences between tasks in frontal theta activity predicted differences in learning rate both for positive, $rho(20) = .43, p = .06$, and negative feedbacks, $rho(19) = .49, p < .05$ (Figure 9). In this later computation, the overall learning rate of the PL was computed as the average of the learning rates obtained in all blocks.

DISCUSSION

In the present work, we studied the modulation of two neurophysiologic components (the FRN and midfrontal theta oscillatory activity) with RPE and learning rate. These two signals have been suggested to be generated in the mPFC, a hypothesis that has been tested by using source modeling and confirmed by intracranial studies (Cohen et al., 2008; Luu et al., 2003, 2004). The participants performed an RVL task in which unpredictable changes of rule occurred and a PL task with multiple blocks of new action–outcome associations without reversal rules. Variations in electrophysiological responses to RPE and learning rate, across and within participants, were analyzed. Two main results were extracted from the data. First, FRN and frontal theta activity were modulated by unsigned prediction error. In addition, variations in frontal theta activity reflected variations in the learning rate across participants and tasks.

The present results show the first evidence that there is a fast evaluation of the learning rate in the mPFC, which is parallel to the processing of expectancy deviations. Three independent results support this claim. First, variations in theta activity across participants were correlated with individual learning rates during the RVL task. Second, theta activity was also sensitive to variations in learning rate within participants across the different blocks of the PL task. Finally, differences in frontal theta activity between the two tasks were predicted by differences in their learning rate. These results provide evidence that frontal theta oscillatory activity is modulated not only on the basis of an unsigned RPE as previously reported (Cavanagh et al., 2011, see also below) but also by the learning rate across and within participants. Learning rate is a key feature of the RL model and controls the impact of new information on the next action value estimate. For example, a learning rate value of 1 indicates that only new acquired information is being considered; in contrast, a learning rate value of 0 shows that new information is not being used, that is, there is no learning from new experience. Therefore, the learning rate determines the weight of the value of RPE to update old estimates (Sutton & Barto, 1998). Previous studies have proposed a relationship between mPFC activity, specifically in ACC, and the learning rate (Jocham et al., 2009; Krugel, Biele, Mohr, Li, & Heekeren, 2009; Behrens et al., 2007; Yoshida & Ishii, 2006; Walton, Devlin, & Rushworth, 2004). For instance, Behrens et al. (2007) showed that in high volatile (fast-changing) environments, the learning rate was higher than in stable environments, and those differences in learning rate resulted in differences in ACC activity. Additionally, and consistent with other studies (Jocham et al., 2009; Krugel et al., 2009), individual differences in learning rate were correlated with ACC BOLD signal. This increase of ACC activity could

parallel the increases of frontal theta activity observed in our study.

The second main finding of the current study is that the FRN ERP and frontocentral theta oscillatory activity are associated to the unsigned RPE of the current trial in the two different experimental paradigms used. These results do not support one of the most influential models about the origin of frontocentral negativities (specially the FRN): the RL theory (Holroyd & Coles, 2002). This model postulates that phasic reduction in the firing of midbrain dopaminergic neuron activity following worse than expected events (Schultz, 1997) is transmitted to ACC, which in turn uses this information to adjust behavior. Some studies have supported this theory by showing that negative feedback elicits greater FRN than positive feedback (Philiastides et al., 2010; Gehring & Willoughby, 2002; Holroyd & Coles, 2002). Similarly, theta activity has also been associated with negative RPEs (Cavanagh, Frank, Klein, & Allen, 2010; Marco-Pallares et al., 2008; Cohen et al., 2007). However, our results would argue against a specific valence effect (whether positive or negative) for both theta activity and FRN amplitude. In contrast, they would agree with recent studies showing that frontocentral theta activity and FRN amplitude also responds to the unsigned (both positive and negative) RPEs (Ferdinand et al., 2012; Cavanagh et al., 2011). In addition, recent findings in nonhuman animal studies have also shown that mPFC neurons are sensitive to surprising outcomes regardless of their valence (Bryden, Johnson, Tobia, Kashtelyan, & Roesch, 2011; Hayden, Heilbronner, Pearson, & Platt, 2011). Present results partially agree with a recent study that tries to dissociate the sensitivity of both FRN amplitudes and theta power increases to outcome valence and probability (Hajihosseini & Holroyd, 2013). The authors showed that, although both FRN and evoked theta power increases were sensitive to outcome valence and probability, they were more strongly determined by outcome valence. In contrast, induced theta power was more affected by outcome probability, reflecting dissociation between FRN amplitude and midfrontal theta power increases. Additionally, these results support the idea that dissociable processes, such as outcome and valence processing, engage simultaneously similar brain mechanism as midfrontal theta oscillatory activity.

The relationship of FRN and theta oscillatory activity with both learning rate and unsigned RPE fits well with a new model that proposes that the mPFC detects action–outcome discrepancies independently from their affective valence (Alexander & Brown, 2011). The predicted response–outcome model (PRO model) is able to correctly simulate some of the previously reported results on the activity of mPFC in error processing, conflict detection, and action monitoring. According to the model, mPFC neurons would fire when an action yields an unexpected outcome, that is, when the outcome is unexpected (positive surprise), but also when an expected outcome does not appear (negative surprise). Therefore, according to the PRO model, both the negative and positive prediction errors in the RL model are unexpected outcomes and therefore unexpected nonoccurrences of the expected response. The modulation of theta and FRN activity with unsigned prediction error would then be related to the surprise signal of the mPFC. In addition, the PRO model also predicts greater activity of the mPFC in environments showing greater variability (the RVL task compared with the PL task) as surprises are more constant in less predictable environments. Therefore, results in Behrens et al. (2007) showing that the mPFC tracks the volatility of the environment as well as present results showing that theta activity are greater in the RVL task than in the PL task would also be explained by the PRO model.

Similar surprise signals have been reported in other brain regions connected to the mPFC such as the amygdala (Paus, 2001) and its major target, the locus coeruleus (Aston-Jones & Cohen, 2005), which is the main noradrenergic nucleus of the brain. Indeed, pharmacological studies have shown that noradrenergic drugs that lead to an increase of noradrenergic release increase FRN amplitude (Riba, Rodríguez-Fornells, Morte, Münte, & Barbanoj, 2005). Thus, FRN amplitude and theta activity could be related to attentional signals transmitted from the locus coeruleus by noradrenergic neurotransmission rather than reflect increase/decreases of dopamine in the ventral tegmental area. However, this requires further research combining different drugs to study the different roles of both dopamine and noradrenalin in outcome monitoring.

The surprise signal reflected by FRN and theta oscillatory activity signal is also consistent with attentional models that suggest that unexpected outcomes may drive learning by increasing attention to subsequent events (Pearce & Hall, 1980). Theta oscillatory activity could then indicate the need to reallocate processing resources as focusing attention on the most relevant information. The idea that the mPFC generates attention-related signals is consistent with a growing body of literature showing the mPFC's role in attention and cognitive control (Shackman et al., 2011; Kerns et al., 2004; Botvinick, Braver, Barch, Carter, & Cohen, 2001). Indeed, theta oscillations are an optimal mechanism of communication between distant brain regions of a same network (Buzsáki & Draguhn, 2004). mPFC is functionally connected to dorsolateral pFC (dlPFC) through theta rhythms (Brázdil et al., 2009), and both structures cooperate to regulate behavior (Botvinick et al., 2001). Therefore, mPFC might monitor internal and external cues to detect unexpected action–outcome discrepancies and recruit the dlPFC according to task demands. Depending on the environmental cues and task needs, mPFC might request different cognitive control adjustments to the dlPFC, such as the increase of more basic information-processing pathways (Kerns, 2006; Kerns et al., 2004; Botvinick et al., 2001) or the engagement of working memory to retain information (Botvinick et al., 2001) to make current context more

relevant than previous experience. Indeed, increases in frontal theta activity have been observed to reflect task difficulty (Gevins, Smith, McEvoy, & Yu, 1997), to increase with memory load in working memory (Deiber et al., 2007; Jensen & Tesche, 2002), and to be related to a wide variety of tasks under situations of conflict and error (Cavanagh et al., 2012).

Acknowledgments

Supported by the Ramon y Cajal program to J. M. P. (RYC-2007-01614), FPI to E. M. H. (BES-2010-032702), Spanish Government grants (PSI2009-09101 and PSI2012-37472 to J. M. P.), and grants from the Catalan Government (2009-SGR-93).

Reprint requests should be sent to Josep Marco-Pallarés, Department of Basic Psychology-IDIBELL, L'Hospitalet de Llobregat, University of Barcelona, Campus Bellvitge, Barcelona 08097, Spain, or via e-mail: josepmarco@gmail.com.

REFERENCES

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

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28,* 403–450.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10,* 1214–1221.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

Brázdil, M., Babiloni, C., Roman, R., Daniel, P., Bares, M., Rektor, I., et al. (2009). Directional functional coupling of cerebral rhythms between anterior cingulate and dorsolateral prefrontal areas during rare stimuli: A directed transfer function analysis of human depth EEG signal. *Human Brain Mapping, 30,* 138–146.

Bryden, D. W., Johnson, E. E., Tobia, S. C., Kashtelyan, V., & Roesch, M. R. (2011). Attention for learning signals in anterior cingulate cortex. *The Journal of Neuroscience, 31,* 18266–18274.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304,* 1926–1929.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2011). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral Cortex, 22,* 2575–2586.

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage, 49,* 3198–3209.

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. B. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology, 49,* 220–238.

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience, 23,* 936–946.

Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage, 35,* 968–978.

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *The Journal of Neuroscience, 27,* 371–378.

Cohen, M. X., Ridderinkhof, K. R., Haupt, S., Elger, C. E., & Fell, J. (2008). Medial frontal cortex and response conflict: Evidence from human intracranial EEG and medial frontal cortex lesion. *Brain Research, 1238,* 127–142.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience, 22,* 4563–4567.

Deiber, M.-P., Missonnier, P., Bertrand, O., Gold, G., Fazio-Costa, L., Ibañez, V., et al. (2007). Distinction between perceptual and attentional processing in working memory tasks: A study of phase-locked and induced oscillatory brain dynamics. *Journal of Cognitive Neuroscience, 19,* 158–172.

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

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295,* 2279–2282.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7,* 374–385.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66,* 585–595.

Hajihosseini, A., & Holroyd, C. B. (2013). Frontal midline theta and N200 amplitude reflect complementary information about expectancy and outcome evaluation. *Psychophysiology, 50,* 550–562.

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise signals in anterior cingulate cortex: Neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *The Journal of Neuroscience, 31,* 4178–4187.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *The European Journal of Neuroscience, 15,* 1395–1399.

Jocham, G., Neumann, J., Klein, T. A., Danielmeier, C., & Ullsperger, M. (2009). Adaptive coding of action values in the human rostral cingulate zone. *The Journal of Neuroscience, 29,* 7489–7496.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–779.

Kerns, J. G. (2006). Anterior cingulate and prefrontal cortex activity in an fMRI study of trial-to-trial adjustments on the Simon task. *Neuroimage, 33,* 399–405.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303,* 1023–1026.

Krugel, L. K., Biele, G., Mohr, P. N. C., Li, S.-C., & Heekeren, H. R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 17951–17956.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science, 14,* 47–53.

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology, 115,* 1821–1835.

Marco-Pallares, J., Cucurell, D., Cunillera, T., García, R., Andrés-Pueyo, A., Münte, T. F., et al. (2008). Human oscillatory activity associated to reward processing in a gambling task. *Neuropsychologia, 46,* 241–248.

Oliveira, F. T. P., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience, 19,* 1994–2004.

Paus, T. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews Neuroscience, 2,* 417–424.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

Philiastides, M. G., Biele, G., Vavatzanidis, N., Kazzer, P., & Heekeren, H. R. (2010). Temporal dynamics of prediction error processing during reward-based decision making. *Neuroimage, 53,* 221–232.

Riba, J., Rodríguez-Fornells, A., Morte, A., Münte, T. F., & Barbanoj, M. J. (2005). Noradrenergic stimulation enhances human action monitoring. *The Journal of Neuroscience, 25,* 4370–4374.

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences, 8,* 410–417.

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current Opinion in Neurobiology, 7,* 191–197.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience, 12,* 154–167.

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Walton, M. E., Croxson, P. L., Behrens, T. E. J., Kennerley, S. W., & Rushworth, M. F. S. (2007). Adaptive decision making and value in the anterior cingulate cortex. *Neuroimage, 36 (Suppl. 2),* T142–T154.

Walton, M. E., Devlin, J. T., & Rushworth, M. F. S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience, 7,* 1259–1265.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning, 8,* 279–292.

Wu, Y., & Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research, 1286,* 114–122.

Yoshida, W., & Ishii, S. (2006). Resolution of uncertainty in prefrontal cortex. *Neuron, 50,* 781–789.