

Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning

Henry W. Chase¹, Rachel Swainson², Lucy Durham¹, Laura Benham¹,
and Roshan Cools^{1,3}

Abstract

■ We assessed electrophysiological activity over the medial frontal cortex (MFC) during outcome-based behavioral adjustment using a probabilistic reversal learning task. During recording, participants were presented two abstract visual patterns on each trial and had to select the stimulus rewarded on 80% of trials and to avoid the stimulus rewarded on 20% of trials. These contingencies were reversed frequently during the experiment. Previous EEG work has revealed feedback-locked electrophysiological responses over the MFC (feedback-related negativity; FRN), which correlate with the negative prediction error [Holroyd, C. B., & Coles, M. G. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109, 679–709, 2002] and which predict outcome-based adjustment of decision values [Cohen, M. X., & Ranganath, C. Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, 27, 371–378,

2007]. Unlike previous paradigms, our paradigm enabled us to disentangle, on the one hand, mechanisms related to the reward prediction error, derived from reinforcement learning (RL) modeling, and on the other hand, mechanisms related to explicit rule-based adjustment of actual behavior. Our results demonstrate greater FRN amplitudes with greater RL model-derived prediction errors. Conversely expected negative outcomes that preceded rule-based behavioral reversal were not accompanied by an FRN. This pattern contrasted remarkably with that of the P3 amplitude, which was significantly greater for expected negative outcomes that preceded rule-based behavioral reversal than for unexpected negative outcomes that did not precede behavioral reversal. These data suggest that the FRN reflects prediction error and associated RL-based adjustment of decision values, whereas the P3 reflects adjustment of behavior on the basis of explicit rules. ■

INTRODUCTION

The medial frontal cortex (MFC) has been implicated in the flexible adjustment of behavior on the basis of changes in reward and punishment values (Cohen & Ranganath, 2007; Rushworth, Buckley, Behrens, Walton, & Bannerman, 2007; Roelofs, van Turenout, & Coles, 2006; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). However, debate continues over its precise contribution. Here we focus on one of its electrophysiological signatures, the feedback-related negativity (FRN), to further elucidate the role of the MFC (Holroyd & Coles, 2008; Gehring & Willoughby, 2002, but see Nieuwenhuis, Slagter, von Geusau, Heslenfeld, & Holroyd, 2005; van Veen, Holroyd, Cohen, Stenger, & Carter, 2004) in distinct forms of outcome-based adjustment.

It has long been accepted that behavior can be adjusted by one or more of multiple decision systems (Daw, Niv, & Dayan, 2005). For example, behavioral adjustment might be controlled by incremental “caching”-like reinforcement learning (RL), associated most commonly with the (dorso-lateral) striatum, or by more explicit rules, associated most

commonly with the pFC. Here we aim to assess the degree to which the FRN, measured over the MFC, reflects outcome-induced adjustment of decisions on the basis of the reward prediction error, derived from a standard RL model, or alternatively outcome-based adjustment of decisions on the basis of an explicit rule, given to participants during task instruction.

This question relates to the ongoing debate about the function of the MFC (Botvinick, 2007; Rushworth et al., 2007; Devinsky, Morrell, & Vogt, 1995). In particular, it speaks to current hypotheses that the MFC signals the need to adjust performance (Roelofs et al., 2006; Ridderinkhof et al., 2004; MacDonald, Cohen, Stenger, & Carter, 2000). Specifically, the present study aims to extend recent findings by Cohen and Ranganath (2007), who reported a relationship between FRN amplitude and behavioral adjustment during a probabilistic decision making task. Critically, in this task, there was no explicit rule, and participants were required to select between two options on the screen, each of which was rewarded with a 50% probability. The authors observed greater FRN amplitude after reward omission than after reward. In addition, FRN amplitude (accompanying feedback on trial n) was larger when the

¹University of Cambridge, ²University of Aberdeen, ³Radboud University Nijmegen, The Netherlands

subject switched their response on the subsequent trial ($n + 1$) to the other option (i.e., not the option they had selected on trial n) compared with when they did not switch. The authors presented a prediction error learning model that accounted for their FRN data. This model registered a larger prediction error accompanying feedback before a response switch than before a response repeat, controlling for feedback valence. The present study aims to extend these findings by assessing FRN amplitudes during unexpected reward omissions that do not lead to actual behavioral adjustment with those during expected reward omissions that do lead to behavioral adjustment.

To this end, we used a paradigm that has been commonly used to study outcome-based behavioral adjustment, that is, the probabilistic reversal learning task (Cools, Clark, Owen, & Robbins, 2002; Swainson et al., 2000). After having obtained a learning criterion, the contingencies reverse and participants adjust their behavior accordingly. In our version of the task, we instructed participants to reverse responding to the previously punished but newly rewarded stimulus only when they were sure that the contingencies had changed. This explicit rule enabled us to separately assess, on the one hand, mechanisms related to the reward prediction error and, on the other hand, mechanisms related to explicit rule-based adjustment of actual behavior. Indeed as instructed, after contingency reversal, participants continued to choose the previously rewarded stimulus until they had acquired sufficient and unambiguous evidence for the need to adjust responding. Accordingly, the negative prediction error was largest during a first reward omission but then reduced with successive reward omissions, until, just before reversal, the subject was no longer surprised, hence reversed responding on the next trial. Thus, if the FRN reflects negative prediction error (and associated RL-based adjustment of decision values), then its amplitude should diminish as reward omissions become better predicted. Indeed, its amplitude should be smallest just before actual behavioral reversal. However, if the FRN reflects explicit rule-based behavioral adjustment, then its amplitude should be largest (more negative) during the reward omission just before behavioral reversal. Consistent with previous theorizing and empirical data (Cohen & Ranganath, 2007; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Holroyd & Coles, 2002; Nieuwenhuis et al., 2002), we anticipated that the FRN amplitude would correlate positively with the reward prediction error and not with rule-based behavioral adjustment.

To test this hypothesis, we assessed the relationship between the FRN amplitude and (both the positive and the negative) the reward prediction error, derived from a standard RL model, not only across different trial types (as has been done previously; Holroyd et al., 2003; Nieuwenhuis et al., 2002) but also by examining trial-by-trial variation in FRN amplitude. Specifically, we used linear regression analysis, with the FRN amplitude as the dependent measure and the model-derived prediction error as the inde-

pendent measure, to test whether the FRN scales with increasing magnitude of prediction error (as reflected in the slope of the regression line).

METHODS

Participants

Thirteen healthy, right-handed students from the University of Cambridge (5 men, age 19–22 years) were recruited. Exclusion criteria were recent head injury, psychiatric or neurological disease, use of psychiatric drugs, and color blindness. Participants received a small fee for participation and provided informed consent approved by the University of Cambridge Research Ethics Committee.

Experimental Design

The study used a probabilistic reversal learning paradigm, adapted from Cools, Lewis, Clark, Barker, and Robbins (2007) and Cools et al. (2002), administered on a desktop computer. Responses were recorded using a button box. On each trial, two Hiragana characters were presented on either side of a central fixation point. Participants were instructed to select one of the characters and received feedback on their choice in the form of either a green smiling face (reward) or a red frowning face (punishment). One of the stimuli (the correct stimulus) would result in reward on 80% of trials, whereas the other (the incorrect stimulus) would be rewarded on 20% of trials. Intermittently, the contingencies reversed, after between 4 and 10 choices of the highly rewarded stimulus. After contingency reversal, selection of the previously correct stimulus would lead to punishment on every trial until the subject selected the now-correct stimulus. Each block contained four reversals. Participants completed 10 blocks, preceded by a short practice session. Participants were presented with the following task instructions.

You will see two pictures. One of the pictures is correct and the other is wrong. You have to choose the correct pattern on each go. On some goes, the computer will tell you that you were wrong even if you chose the correct pattern. Your task is to stick to the pattern that is usually correct. Sometimes the rule may change so that the other pattern is now usually correct. You then have to follow this new rule and choose the new pattern so that, in general, you still get as many green smiley faces as possible over the whole of the task. It is important that you only start choosing the other pattern when you are sure that the rule has changed!!! The rule will change several times, but there is no way of predicting when it will change. The same two patterns will be presented throughout the task. Try to respond as quickly as you can. If you respond too slowly, then the computer will tell you that you were "too late." Try to avoid this as much as possible. Fixate on the cross in between trials.

Stimuli remained on the screen until the subject made a response on the button box, then, after a 1000-msec delay, feedback was presented for 500 msec. The feedback/subsequent stimuli delay was jittered between 750 and 1250 msec to ensure that feedback-related activity was not confounded by presentation of the next stimulus.

Electrophysiological Recording and ERP Extraction

Scalp electrical activity was recorded with a 128-electrode HydroCel Geodesic Sensor Net (Tucker, 1993). Each electrode was adjusted until its impedence was below 50 k Ω . Data were recorded at 250 Hz, using the vertex electrode (129th electrode) as the reference. The amplifier band-pass was 0.1–100 Hz, and the data were low-pass filtered at 40 Hz off-line. The data were average referenced and corrected for the polar average reference effect (Junghofer, Elbert, Tucker, & Braun, 1999). Epochs of 700 msec (200 msec baseline before feedback presentation, 500 msec after) were extracted. Epochs were excluded if they had amplitudes greater than 70 μ V or a channel variance of 0, as these were likely to contain eye blinks or other artifacts. Channels for which 15% or more of segments contained artifacts were marked as bad, and their data were excluded. In epochs where fewer than 15% channels were marked bad, data from bad channels were interpolated from the remaining channels. Notably only one channel of one epoch for one subject was marked bad on the basis of the electrodes included in the ROIs.

Feedback-locked ERPs were extracted from five trial types to investigate our a priori hypotheses:

1. “valid positive” (vP) feedback coinciding with correct responses (excluding the first correct response after behavioral switching);
2. “spurious negative” (sN) feedback coinciding with correct responses;
3. negative feedback coinciding with erroneous responses after contingency reversal, which were not followed by behavioral switching (“preceding negative”; pN);
4. negative feedback coinciding with the final erroneous response before behavioral switching (“final negative”; fN); and
5. positive feedback coinciding with the first correct response after behavioral switching (“first positive”; fP).

The data from the epochs were averaged across each condition to produce an average waveform for each trial type and each subject. These waveforms were then baseline corrected by subtracting the mean of the first (pre-stimulus) 200 msec from the rest of the epoch.

Feedback-related Negativity

The amplitude of the FRN for each subject was determined by subtracting the average of the preceding and

following positive peaks (Yeung & Sanfey, 2004) from the lowest point within a window 248–296 msec post-feedback (Yeung, Holroyd, & Cohen, 2005), using data from a symmetrical cluster of eight central electrodes (6, 7, 13, 31, 55, 80, 106, and 112; see Figure 1, top marked cluster). If the lowest point was at the edge of the window, the size of the window was widened until the lowest point was then no longer on the edge of the window in order that the nadir of the FRN be identified correctly. The equivalent was performed for the preceding and following peaks (if the highest point was at the edge of the window, the window was increased in size) to ensure that the zeniths were identified. The window for the preceding peak was between 180 and 208 msec postfeedback, and the window for the following peak was between 346 and 376 msec postfeedback.

P3

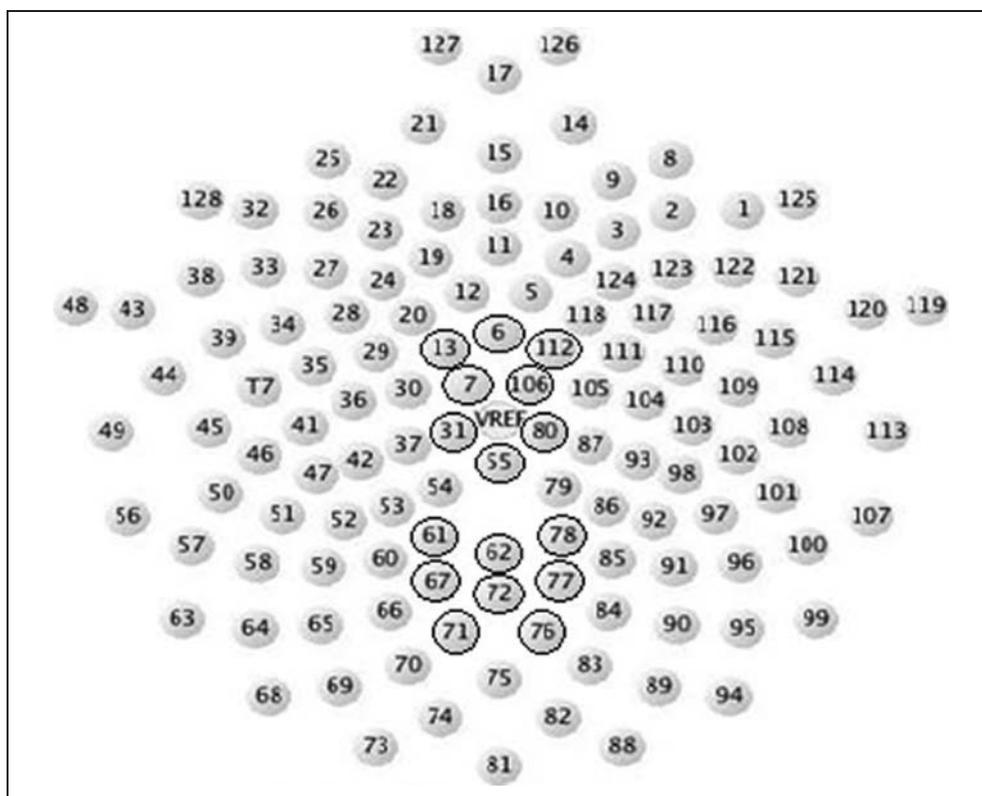
Inspection of the data revealed large P3 responses during selective trial types. We decided to also quantify the P3 amplitude for the following two reasons. First, the measurement of the FRN can be affected by the P3 amplitude. Second, evidence indicates that the P3 can be elicited by detection of changes in task contingency (Barcelo, Escera, Corral, & Perianez, 2006; Johnson & Donchin, 1982) as well as by unpredicted (rewarding) events (Bellebaum & Daum, 2008; Hajcak, Moser, Holroyd, & Simons, 2007; Hajcak, Holroyd, Moser, & Simons, 2005; Hajcak, Moser, Yeung, & Simons, 2005). Our paradigm provides a unique opportunity to disentangle the role of the P3 in the detection of unexpected events, contingency changes, and actual behavioral adaptation. Hence, we used a method similar to that used by Yeung and Sanfey (2004) to obtain independent estimates of FRN and P3 amplitude and applied the trial-by-trial linear regression method to the P3 as well as the FRN. Critically, we performed supplementary analyses to disentangle the estimates of these two ERPs (see Supplementary ERP analyses section).

P3 amplitude was quantified by extracting data from the central as well as a parieto-occipital region (a symmetrical cluster of eight posterior electrodes: 61, 62, 67, 71, 72, 76, 77, and 78; see Figure 1, bottom marked cluster). Because the data in the time range of the P3 component included some noise such that a clear P3 peak was not easy to determine, a measure of mean amplitude over a 300- to 500-msec postfeedback window was used. Of the sensors used, five specific locations correspond closely to locations within the 10–10 system, that is, 6 (FCz), 55 (CpZ), 67 (PO3), 72 (POz), and 77 (PO4) (Luu & Ferree, 2000).

Reinforcement Learning Model

To further investigate our predictions, specifically those inspired by Holroyd and Coles (2002), the RL (*Q* value) model used by Cohen and Ranganath (2007) was implemented.

Figure 1. Diagram showing the distribution of scalp electrodes (anterior electrodes are at the top). Electrodes 6, 7, 13, 31, 55, 80, 106, and 112 were used to acquire central electrical activity (for FRN extraction—black outline, top cluster), whereas electrodes 61, 62, 67, 71, 72, 76, 77, and 78 were used to acquire parietal electrical activity (for P3 extraction—black outline, bottom cluster).



The Q value of the selected stimulus A is updated with new information using the following algorithm:

$$Q_A(t + 1) = Q_A(t) + \alpha\delta$$

where δ is the prediction error ($\text{outcome}(t) - Q_A(t)$), α is the learning rate, t is current trial, and outcome is 1 or -1 depending on feedback valence. Note that the discounting parameter present in Cohen and Ranganath's model was omitted because of a concern that the parameter would reflect the autocorrelation of correct responses as a result of the contingencies used in the task and be greater than 1 if freely estimated. It was therefore set equal to 1.

The relative Q values of stimulus A and B were used to compute the probability of selecting one of the stimuli using the following equation:

$$P_A(t) = \exp(Q_A(t)) / (\exp(Q_A(t)) + \exp(Q_B(t)))$$

Individual learning rate parameters (i.e., α) were determined for each subject by optimizing the fit of the model, that is, by maximizing the model's estimation of the probability of selecting the outcome that the subject picks on each trial using the nonlinear, unconstrained Nelder–Mead simplex method implemented in Matlab 6.5 (MathWorks, Natick, MA). The optimization algorithm was run from different starting points to ensure that the presence of local minima did not influence accurate estimation of the parameter.

Having determined learning rate parameters for each subject, values representing the prediction error δ on

each trial could be determined. The relationship between these values and the FRN and P3 amplitude for each trial were determined separately for each subject. First, all trials were independently baseline corrected. The amplitude of the FRN was determined for each trial using a similar base/peak method to that described earlier, except that the same latency for the FRN dip, preceding peak, and following peak was used for each trial for a given subject. The latencies at the dip and peaks were estimated by averaging all trials for each subject and then finding the smallest voltage in the 248- to 296-msec window and the preceding and following peak voltages. FRN amplitude for each trial was then determined by subtracting the average of the voltages at the peaks from the voltage at the dip. P3 amplitude on every trial was estimated by finding the mean amplitude in the 300- to 500-msec postfeedback window from the parietal electrodes. Using these values, a linear model could then be fitted for each subject:

$$\text{ERP Amplitude} = k\delta + \text{intercept} + \text{noise}$$

Slopes (k) and intercepts were determined using FRN and P3 amplitudes as the dependent measures. We then determined whether the value of the slope was different overall from 0 for the group for each dependent 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 ERP amplitude on individual trials. Separate analyses for positive and negative prediction errors were performed given current controversy

regarding the valence-specificity of the FRN. Although the FRN is commonly thought to accompany outcomes that are worse than expected, recent evidence indicates that it might also have a larger (more negative) amplitude when the outcome is better than expected (Oliveira, McDonald, & Goodman, 2007). If the FRN scales with both types of prediction error, then we would see significant linear relationships (as evidenced by regression slopes that are significantly greater than zero) between FRN and both negative and positive prediction errors. Accordingly, we performed two linear regression analyses, separately considering trials on which prediction error was negative (the outcome was worse than expected) and trials on which the prediction error was positive (outcome better than expected). We then compared the slopes and intercepts of these relationships. If the FRN scales with the magnitude of the negative prediction error (Holroyd & Coles, 2002), then the slope of this relationship should be positive (more negative prediction errors with more negative FRN). If it also scales with the magnitude of the positive prediction error, then the slope of the relationship between the positive prediction error and the FRN should be negative (more positive prediction errors with more negative FRN). In addition, if the prediction error (negative or positive) is the critical determinant of the FRN amplitude, then the intercepts of the two regression equations should be similar. Alternatively, if the FRN amplitude does not scale with increasing negative prediction error but simply reflects a binary evaluation of whether the feedback was positive or negative, then there should be a difference between the intercepts of the two regression equations, but not their slopes (which should be close to zero).

Behavioral and ERP Data Analysis

Parameters of interest for each subject were as follows (see previous paragraphs): mean number of valid positive (vP) trials, spurious negative (sN) trials, preceding negative (pN) trials, final negative (fN) trials, and first positive (fP) trials; mean number of perseverative errors (pN + fN trials); and probability of switching following spurious negative (sN) trials. Analysis of the FRN and P3 amplitude was performed using repeated measures ANOVA, contrasting each of the five trial types. Paired two-tailed *t* tests were used to further assess planned comparisons of primary interest, as outlined in the Introduction. An alpha level of .05 was used in all statistical comparisons. The Greenhouse–Geisser correction was applied when the sphericity assumption was violated.

RESULTS

Behavioral Data

All participants completed 10 blocks, aside from one subject who completed 13 blocks and another who completed 7 blocks because of a computer error. Mean numbers of

trial types were as follows: 322.9 vP trials ($SD = 48.0$), 93.0 sN trials ($SD = 15.0$), 52.6 pN trials ($SD = 17.0$), 39.3 fN ($SD = 5.8$), and 39.3 fP trials ($SD = 5.8$). There were a mean of 2.3 perseverative errors (pN and fN trials) per reversal ($SD = 0.3$), whereas the mean probability of switching cue after a spurious negative trial was 0.075 ($SD = 0.05$).

After artifact detection and removal, the mean numbers of artifact-free epochs were as follows: 239.9 vP trials ($SD = 88.6$), 72.9 sN trials ($SD = 22.4$, range = 26–107), 41.8 pN trials ($SD = 20.3$), 28.5 fN trials ($SD = 10.5$, range = 8–40), and 27.6 fP trials ($SD = 10.5$).

ERP Analysis: FRN

FRN amplitude for each trial type, computed using a peak to peak method, was analyzed using a repeated measures ANOVA with the within-subject factor trial type (five levels). There was a main effect of trial type, $F(2.2, 26.8) = 10.7$, $p < .001$. Paired *t* tests were used to investigate this effect (Figure 2): As predicted, FRN amplitude was largest on spurious negative trials, then preceding negative trials, whereas FRN amplitude was indistinguishable between final negative trials, valid positive trials, and first positive trials. Temporal dynamics of the FRN are displayed in Figure 3A.

Thus, the events on which negative feedback was least expected, that is, the spurious negative (sN) trials and the preceding negative (pN) trials evoked the largest FRN amplitudes. However, the event that led to behavioral switching (fN) was not accompanied by an FRN relative to positive feedback (either vP or fP trials). Hence, the FRN amplitude was large during unexpected punishment events that were not followed by behavioral switching. Conversely, there

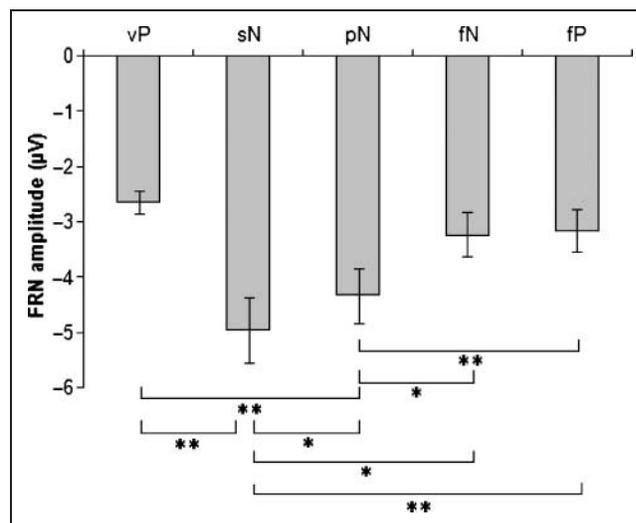


Figure 2. Mean FRN amplitude for each trial type: vP = valid positive; sN = spurious negative; pN = preceding negative; fN = final negative; fP = first positive (* $p < .05$; ** $p < .005$).

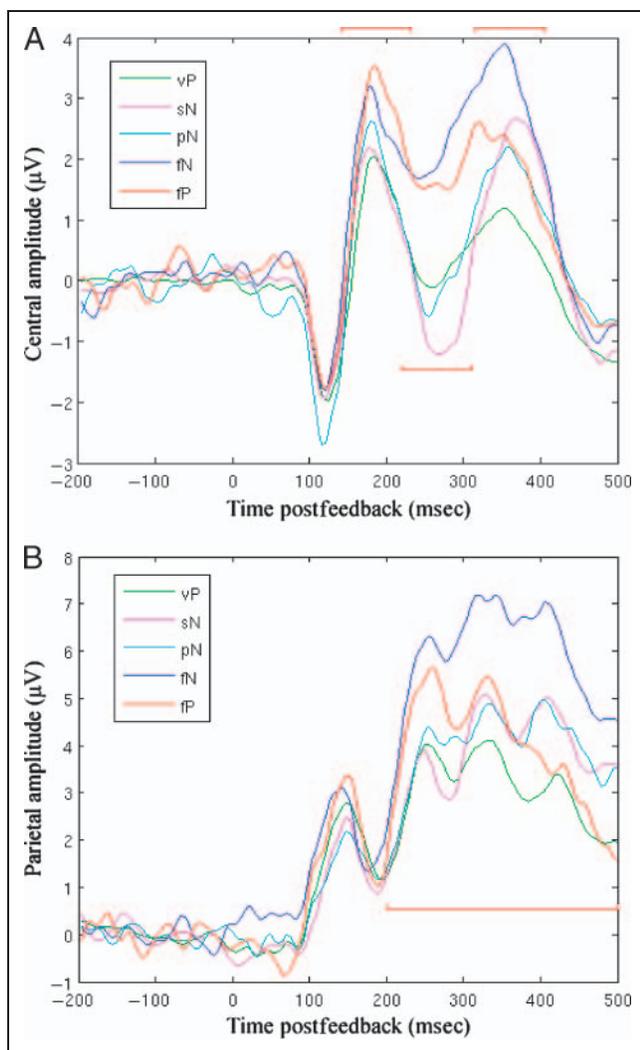


Figure 3. (A) Waveform plots for each trial type at central electrodes, averaged across all participants: vP = valid positive; sN = spurious negative; pN = preceding negative; fN = final negative; fP = first positive. The red markers represent the approximate locations of the peaks and dip used to calculate the FRN (FRN = dip—average of the peaks). (B) Waveform plots for each trial type at parietal electrodes, averaged across all participants. Red marker denotes time window in which the P3 amplitude was determined.

was no FRN on expected punishment events that were followed by behavioral switching.

ERP Analysis: P3

Mean amplitudes were extracted for each trial type from parietal electrodes in a window of 300–500 msec post-feedback and inserted in a repeated measures ANOVA. There was a main effect of trial type, $F(2.2, 26.6) = 12.527$, $p < .001$. Subsequent paired t tests (Figure 4) showed that the largest P3 was produced on final negative (fN) trials. The P3 on these fN trials was significantly larger than that on spurious negative (sN) trials, $t(12) = 4.15$, $p = .001$. The P3 amplitude did not differ significantly between the

first positive (fP), spurious negative (sN), and preceding negative (pN) trials but was larger on each of these negative feedback trials than on the valid positive (vP) trials (although the difference between fP and vP was marginally significant), $t(12) = 1.85$, $p = .09$. Temporal dynamics of the P3 are displayed in Figure 3B.

Thus, P3 amplitude was greater during the negative feedback trials that directly preceded behavioral switching than during the spurious and preceding negative trials. This pattern contrasts with that seen for the FRN.

Supplementary ERP Analyses

One might argue that the difference between the FRN amplitude on sN trials and that on fN trials is confounded by the difference in the P3 amplitude between these trials, which casts doubt on the validity of the contrast. However, we performed a secondary analysis to demonstrate that the same pattern of FRNs was obtained when we matched a subset of sN and fN trials for parietal P3 amplitude. We identified pairs of fN and sN trials whose absolute difference in P3 amplitude (recorded from parietal electrodes) was as small as possible. If two or more fN trials were close to a single sN trial, the fN/sN match with the smallest absolute difference was included, and the fN trial with a worse match was excluded. Amplitudes recorded from central electrodes were compared from this subset of trials (mean number of trials per participant = 22.3, range = 6–33). The maximal difference between those trial types was most clearly evident in the time window in which the FRN is expected (see Figure 5). Mean amplitude of the selected sN and fN trials was approximately matched in the windows that had been used to define the peaks that preceded and followed the dip, first peak—180 to 208 msec, $t(12) = 1.8$, $p = .098$, second peak—356 to

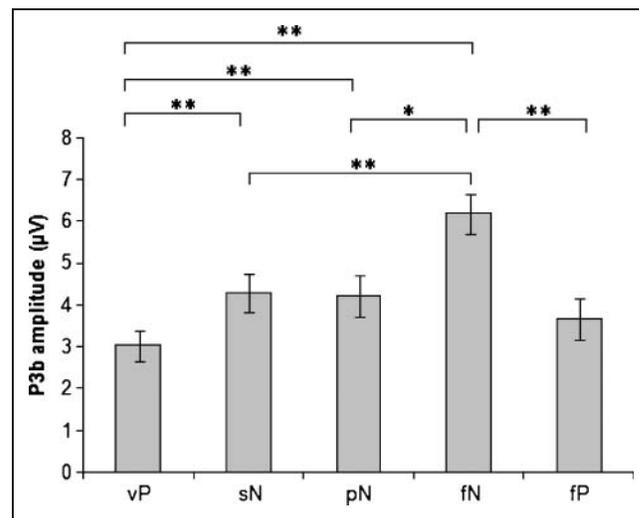


Figure 4. Mean P3 amplitude for each trial type: vP = valid positive; sN = spurious negative; pN = preceding negative; fN = final negative; fP = first positive (* $p < .05$; ** $p < .005$).

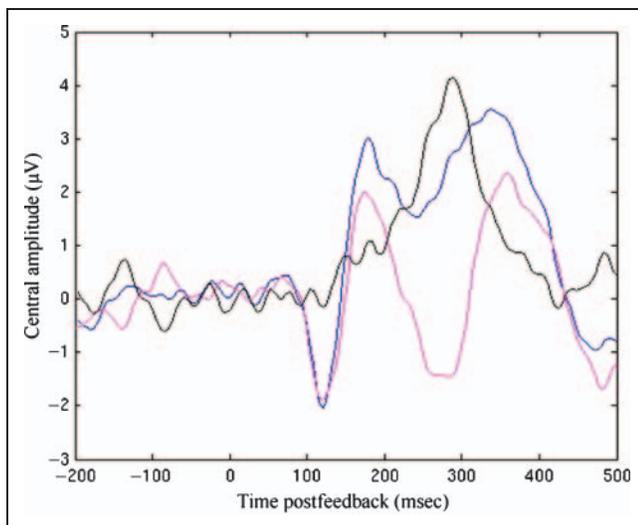


Figure 5. Graph describing the supplementary analysis of fN (blue line) and sN (magenta line) trials using the P3 matching procedure. The difference wave (black line) clearly demonstrates that the maximal difference between the two waveforms is in the time window in which the FRN is expected; hence, the statistical differences between the conditions are not caused by confounding with P3 magnitude.

376 msec, $t(12) = 1.4$, $p = .19$, whereas mean amplitudes were clearly different between the two trial types in the window used to define the dip, 248 to 296 msec, $t(12) = 4.0$, $p = .002$. Central P3 amplitudes in the selected subset were also matched, mean amplitude between 300 and 500 msec, $t(12) = 1.7$, $p = .113$. A similar pattern of data, although with weaker effect sizes, was observed when fN and pN trials were compared in the same way.

Reinforcement Learning Analysis

A simple RL model was fitted to subject's behavioral performance by adjusting the magnitude of the learning rate parameter. The model gave a good approximation of participants' behavioral performance (pseudo- $R^2 = .39$, $SE = 0.01$). Participants had a mean learning rate of 0.73 ($SE = 0.01$). Mean prediction error was determined for each trial type, and there was a significant main effect of trial type, $F(4, 48) = 2278.9$, which was due to highly significant paired t tests for comparisons between all five trial types with each other, $t(12) > 7.3$, $p < .001$ in all cases. As predicted, spurious negative trials were accompanied by the most negative prediction error, followed by preceding negative, then final negative then valid positive, and then first positive trials (see Figure 6).

The relationship between the FRN amplitude and the magnitude of the prediction error was determined on a trial-by-trial basis. As predicted, there was a positive relationship between negative prediction errors and FRN amplitude in all but one of the participants: Greater negative prediction errors were associated with larger (more negative) FRN (Figures 7 and 8). The mean slope for the group was significantly greater than 0, $t(12) = 3.5$, $p = .004$.

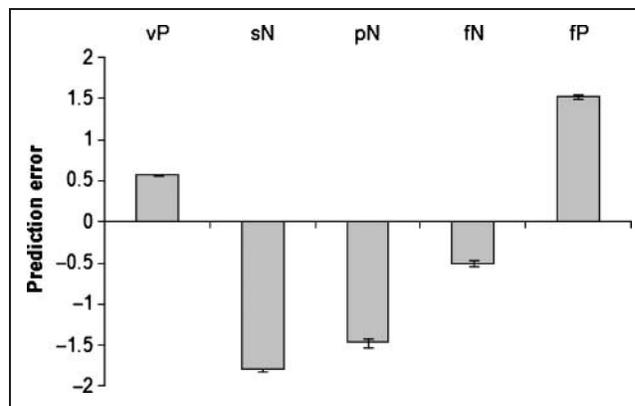


Figure 6. Reinforcement learning model-derived prediction error associated with each trial type: vP = valid positive; sN = spurious negative; pN = preceding negative; fN = final negative; fP = first positive.

There was also a trend toward a significant negative relationship between the positive prediction error and the FRN amplitude (greater positive prediction errors resulting in a larger—more negative—FRN; Figure 7), although the fitted slope was only marginally significantly different from 0, $t(12) = -2.0$, $p = .07$. Paired t tests showed that the slopes of the relationship between the prediction error and the FRN differed significantly between positive and negative prediction errors, $t(12) = 4.6$, $p = .001$, and there was no significant difference between the absolute slopes, $t(12) = 1.6$; $p = .14$. Together, these observations suggest that both negative and positive prediction error scaled linearly with FRN amplitude, although we note that the relationship with the positive prediction error should be treated with caution for several reasons (see Discussion).

There was no significant difference in terms of the estimated intercepts, $t(12) < 1$ (Figure 9). Therefore, this analysis further strengthened the hypothesis that FRN amplitude is sensitive to the magnitude of prediction error

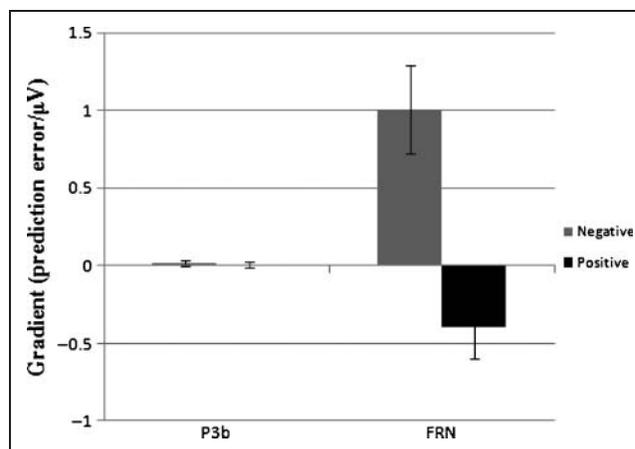
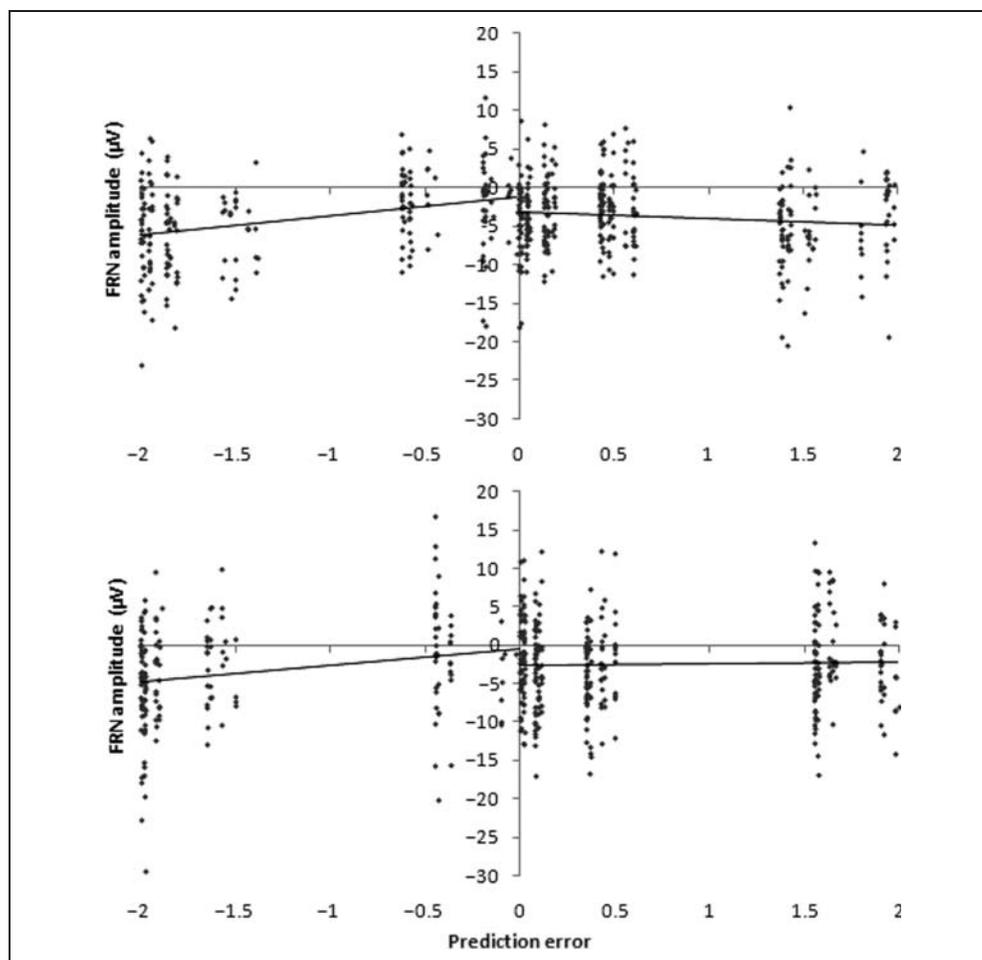


Figure 7. Mean slope for relationships between negative prediction error (gray) or positive prediction error (black) and maximum P3 amplitude (P3) and FRN amplitude (FRN), respectively.

Figure 8. Figure showing the relationship between the FRN amplitude and the prediction error in two representative participants (Participants 4 [top] and 5 [bottom]). Note the positive gradient between prediction error and FRN amplitude for negative prediction errors. Prediction error values are clustered because of a combination of a high learning rate and the relatively stereotyped feedback sequence combinations due to the 80%/20% contingencies.



rather than simply its valence. Finally, the P3 amplitude was fitted to positive and negative prediction errors separately: Fitted slopes were not significantly different from 0 in any of the two models, negative, $t(12) = 1.0, p = .3$, and

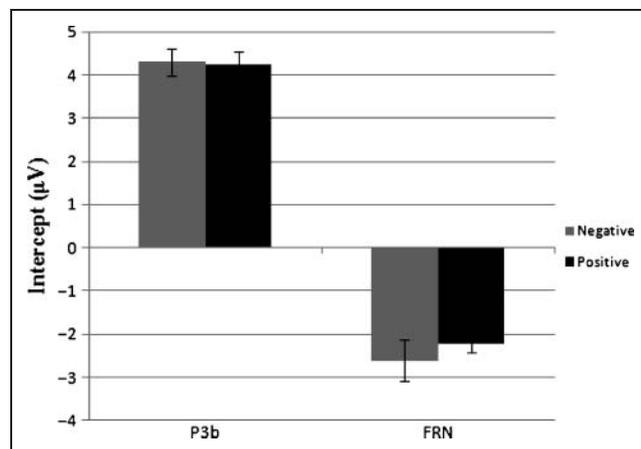


Figure 9. Mean intercept for relationships negative prediction error (gray) or positive prediction error (black) and maximum P3 amplitude (P3) and FRN amplitude (FRN).

positive, $t(12) < 1$, neither were the intercepts significantly different between the two models, $t(12) < 1$.

Summary

Consistent with previous theorizing and empirical data, the FRN amplitude in our study correlated positively with negative prediction error. Specifically, the slope of the relationship between the FRN amplitude and the negative prediction error was significantly different from 0. By contrast, no evidence was found supporting the hypothesis that the FRN amplitude reflects explicit rule-based behavioral adjustment. Instead, the P3 amplitude was a better predictor of rule-based behavioral adjustment because it was largest on the negative feedback trials directly preceding behavioral switching.

DISCUSSION

Adequate behavioral adjustment during probabilistic reversal learning not only depends on RL but might also implicate explicit higher order knowledge. In the task

used here, participants were given such explicit higher order knowledge, that is, the rule to switch responding only when they were sure that the contingencies had changed. This explicit rule enabled us to separately assess, on the one hand, mechanisms related to RL and reward prediction error, which likely contributed to adjustments of covert decision values, and on the other hand, mechanisms related to perhaps more explicit rule-based adjustment of actual overt behavior. Previous data have shown that the FRN predicts overt behavioral adjustment, when such adjustment is accompanied by a large reward prediction error (Cohen & Ranganath, 2007). Conversely, our finding indicates that outcome-based behavioral adjustment is accompanied by a P3 rather than an FRN, when such behavioral adjustment is instead triggered by an explicit rule. Thus, increases in the FRN with behavioral adjustment are seen only if it is triggered by a large prediction error rather than by an explicit rule. Indeed in the study by Cohen and Ranganath (2007), reward probability was 50%, and the adoption of explicit rules or strategies was discouraged. Together these data indicate that different forms of outcome-based behavioral adjustment have distinct electrophysiological signatures, corresponding to the FRN and the P3, respectively.

These data speak to the wider literature in which the observation that distinct neural systems contribute to different forms of behavioral control is receiving an increasing amount of attention (Dayan, 2007; Frank & Claus, 2006; Daw et al., 2005; Holroyd & Coles, 2002), with one system implicated primarily in incremental and integrative RL and another in more flexible and faster “tree-based search” adjustments. The finding that the FRN correlates with the reward prediction error rather than with rule-based adjustment also concurs more generally with the conclusion that the MFC might play an important role in the adjustment of decision values based on the integration of events across reinforcement history (Jocham, Neumann, Klein, Danielmeier, & Ullsperger, 2009; Holroyd & Coles, 2008; Behrens, Woolrich, Walton, & Rushworth, 2007). The finding that, by contrast, the P3 was largest just before (and just after) rule-based behavioral adjustment concurs with previous data supporting a relationship between the P3 amplitude and the detection and (cued) implementation of changes in task contingencies (Fleming, Mars, Gladwin, & Haggard, 2009; Barcelo et al., 2006; Donchin & Coles, 1988; Johnson & Donchin, 1982). However, there was no clear evidence for an association between P3 amplitude and prediction error or valence (Bellebaum & Daum, 2008; Hajcak et al., 2007; Hajcak, Holroyd, et al., 2005; Hajcak, Moser, et al., 2005). Likewise, accounts of P3 amplitude (Duncan-Johnson & Donchin, 1977; Squires, Wickens, Squires, & Donchin, 1976), which focus on the unpredictability of the sequence of different types of event (in this case, positive or negative feedback), would not easily capture this pattern of data because, in general, the longer the sequence of a particular type of trial, the smaller the P3 amplitude. The probabilistic reversal learning task is somewhat un-

usual in that the longer the sequence of negative feedback, the greater the likelihood of a change in task contingency: In many tasks, it is the surprising trial itself that can be seen to signal a change in the local stimulus probability (Mars et al., 2008). One possible explanation for the apparent discrepancy with previous data is that, unlike in previous studies, the most unexpected outcome in our task is not necessarily also the most behaviorally relevant or salient. Thus, the P3 might reflect an aspect of behavioral relevance not dependent on the violation of a stimulus-outcome expectancy and one which is likely to involve the updating of stimulus-response associations. In the present paradigm, this is better coupled to the detection of a change in the rule rather than a change in the associative strength of the stimulus itself (which is reflected by the FRN).

There are a number of open questions regarding the neural systems that mediate these distinct electrophysiological correlates of probabilistic reversal learning. In particular, studies with lesion patients (e.g., Barcelo & Knight, 2007) may elucidate the necessary contribution of the regions activated during the distinct trial types (as revealed by previous fMRI work with the paradigm; Cools et al., 2002) both to FRN and P3 generation and to behavior. Candidate regions include not only the MFC and the parietal cortex but also the ventral striatum and the ventrolateral PFC (Cools et al., 2002).

Likewise, questions regarding the influence of neuromodulators on these processes require further study. For example, according to the model of Holroyd and Coles (2002), dopamine might be critically implicated in the generation of the FRN. Specifically, a phasic reduction in the firing of dopamine neurons could disinhibit layer V neurons in the MFC, allowing these cells to become synchronously depolarized. Our finding that the FRN amplitude tended to correlate positively, albeit only marginally with the positive prediction error, provides a challenge to this view. Indeed unexpected positive events are generally accepted to be accompanied by bursts rather than dips in the firing of dopamine neurons (Hollerman & Schultz, 1998; but see Brischox, Chakraborty, Brierley, & Ungless, 2009; Matsumoto & Hikosaka, 2009). By contrast, our data are compatible with other reports showing that the FRN amplitude is greater when outcome expectations are violated, regardless of the expected valence of the outcome (Oliveira et al., 2007). Nonetheless, caution is warranted when interpreting this finding for three reasons. First, the correlation did not quite reach statistical significance. Second, our *Q*-learning model is not optimized for capturing the dynamics of any positive prediction error in our higher order reversal task (Hampton, Bossaerts, & O’Doherty, 2006). For example, participants are unlikely to be surprised when they receive reward after a contingency reversal, yet this FP trial is coded as being accompanied by a high positive prediction error (Figure 6). Third, most importantly, it should be noted that our base/peak method of evaluating the FRN is biased against detection of positivities observed within the 248- to 296-msec window as we determined the

lowest point within this window. Hence, the magnitude of any positive deflection within this period would have been poorly estimated by our measurement of the base amplitude. Future studies should further elucidate the valence-specificity of the FRN.

There is evidence that dopaminergic and serotonergic manipulations influence human probabilistic reversal learning, in terms of both behavioral performance (Chamberlain et al., 2006; Cools, Barker, Sahakian, & Robbins, 2001; Mehta, Swainson, Ogilvie, Sahakian, & Robbins, 2001) and neural correlates (Cools et al., 2007; Evers et al., 2005). Specifically, the dopamine-enhancing drugs L-dopa and methylphenidate were recently observed to modulate BOLD activity in the ventral striatum during final reversal errors but not during spurious negative feedback in this task (Dodds et al., 2008; Cools et al., 2007). Conversely, serotonergic manipulation by the dietary tryptophan depletion procedure modulated activity in the MFC, and this effect was not restricted to the final reversal errors but extended to the spurious negative feedback events (Cools, Roberts, & Robbins, 2008; Evers et al., 2005). Future study should address the obvious next question, that is, whether the switch-specific P3 and the unexpected feedback-related FRN are differentially modulated by dopaminergic and serotonergic manipulations, respectively. In addition, noradrenergic mechanisms might also influence the amplitude of the P3 (Nieuwenhuis, Aston-Jones, & Cohen, 2005) and FRN (Riba, Rodriguez-Fornells, Morte, Munte, & Barbanj, 2005).

Acknowledgments

The work was conducted within the Behavioral and Clinical Neuroscience Institute at the University of Cambridge, cofunded by the Medical Research Council and the Wellcome Trust to Prof. Trevor W. Robbins, Prof. Barbara J. Sahakian, Prof. Barry J. Everitt, and Dr. Angela C. Roberts. R. C. was supported by a Royal Society University Research Fellowship. H. W. C. was supported by a 3-year MRC studentship while this work was carried out. We thank Rudolph Cardinal and anonymous reviewers for helpful comments on the manuscript.

Reprint requests should be sent to Henry W. Chase, School of Psychology, University Park, University of Nottingham, Nottingham NG7 2RD, UK, or via e-mail: henry.chase@nottingham.ac.uk.

REFERENCES

Barcelo, F., Escera, C., Corral, M. J., & Perianez, J. A. (2006). Task switching and novelty processing activate a common neural network for cognitive control. *Journal of Cognitive Neuroscience*, *18*, 1734–1748.

Barcelo, F., & Knight, R. T. (2007). An information-theoretical approach to contextual processing in the human brain: Evidence from prefrontal lesions. *Cerebral Cortex*, *17*(Suppl. 1), i51–i60.

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

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, *27*, 1823–1835.

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, *7*, 356–366.

Brischoux, F., Chakraborty, S., Brierley, D. I., & Ungless, M. A. (2009). Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proceedings of the National Academy of Sciences, U.S.A.*, *106*, 4894–4899.

Chamberlain, S. R., Muller, U., Blackwell, A. D., Clark, L., Robbins, T. W., & Sahakian, B. J. (2006). Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science*, *311*, 861–863.

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

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral Cortex*, *11*, 1136–1143.

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. *Journal of Neuroscience*, *22*, 4563–4567.

Cools, R., Lewis, S. J., Clark, L., Barker, R. A., & Robbins, T. W. (2007). L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. *Neuropsychopharmacology*, *32*, 180–189.

Cools, R., Roberts, A. C., & Robbins, T. W. (2008). Serotonergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences*, *12*, 31–40.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*, 1704–1711.

Dayan, P. (2007). Bilinearity, rules, and prefrontal cortex. *Frontiers in Computational Neuroscience*, *1*, 1.

Devinsky, O., Morrell, M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain*, *118*, 279–306.

Dodds, C. M., Muller, U., Clark, L., van Loon, A., Cools, R., & Robbins, T. W. (2008). Methylphenidate has differential effects on blood oxygenation level-dependent signal related to cognitive subprocesses of reversal learning. *Journal of Neuroscience*, *28*, 5976–5982.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating. *Behavioural and Brain Sciences*, *11*, 357–374.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456–467.

Evers, E. A., Cools, R., Clark, L., van der Veen, F. M., Jolles, J., Sahakian, B. J., et al. (2005). Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. *Neuropsychopharmacology*, *30*, 1138–1147.

Fleming, S. M., Mars, R. B., Gladwin, T. E., & Haggard, P. (2009). When the brain changes its mind: Flexibility of action selection in instructed and free choices. *Cerebral Cortex*, *19*, 2352–2360.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making and reversal. *Psychological Review*, *113*, 300–326.

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

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, *42*, 161–170.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity

- and violations of reward prediction in gambling tasks. *Psychophysiology*, *44*, 905–912.
- Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*, 151–160.
- Hampton, A. N., Bossaerts, P., & O’Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience*, *26*, 8360–8367.
- Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, *1*, 304–309.
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.
- Holroyd, C. B., & Coles, M. G. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior. *Cortex*, *44*, 548–559.
- Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *NeuroReport*, *14*, 2481–2484.
- Jocham, G., Neumann, J., Klein, T. A., Danielmeier, C., & Ullsperger, M. (2009). Adaptive coding of action values in the human rostral cingulate zone. *Journal of Neuroscience*, *29*, 7489–7496.
- Johnson, R., Jr., & Donchin, E. (1982). Sequential expectancies and decision making in a changing environment: An electrophysiological approach. *Psychophysiology*, *19*, 183–200.
- Junghofer, M., Elbert, T., Tucker, D. M., & Braun, C. (1999). The polar average reference effect: A bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology*, *110*, 1149–1155.
- Luu, P., & Ferree, T. (2000). *Determination of the geodesic sensor nets’ average electrode positions and their 10-10 international equivalents* (pp. 1–11). Eugene, OR: Electrical Geodesics, Inc.
- MacDonald, A. W., III, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1838.
- Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., et al. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience*, *28*, 12539–12545.
- Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*, 837–841.
- Mehta, M. A., Swainson, R., Ogilvie, A. D., Sahakian, B. J., & Robbins, T. W. (2001). Improved short-term spatial memory but impaired reversal learning following the dopamine D(2) agonist bromocriptine in human volunteers. *Psychopharmacology (Berlin)*, *159*, 10–20.
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol Bull*, *131*, 510–532.
- Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G., Holroyd, C. B., Kok, A., et al. (2002). A computational account of altered error processing in older age: Dopamine and the error-related negativity. *Cognitive, Affective & Behavioral Neuroscience*, *2*, 19–36.
- Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J., Heslenfeld, D. J., & Holroyd, C. B. (2005). Knowing good from bad: Differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience*, *21*, 3161–3168.
- Oliveira, F. T., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience*, *19*, 1994–2004.
- Riba, J., Rodriguez-Fornells, A., Morte, A., Munte, T. F., & Barbanj, M. J. (2005). Noradrenergic stimulation enhances human action monitoring. *Journal of Neuroscience*, *25*, 4370–4374.
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447.
- Roelofs, A., van Turenout, M., & Coles, M. G. (2006). Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 13884–13889.
- Rushworth, M. F., Buckley, M. J., Behrens, T. E., Walton, M. E., & Bannerman, D. M. (2007). Functional organization of the medial frontal cortex. *Current Opinion in Neurobiology*, *17*, 220–227.
- Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*, *193*, 1142–1146.
- Swainson, R., Rogers, R. D., Sahakian, B. J., Summers, B. A., Polkey, C. E., & Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson’s disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia*, *38*, 596–612.
- Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, *87*, 154–163.
- van Veen, V., Holroyd, C. B., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2004). Errors without conflict: Implications for performance monitoring theories of anterior cingulate cortex. *Brain and Cognition*, *56*, 267–276.
- Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex*, *15*, 535–544.
- Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience*, *24*, 6258–6264.