CORTEX 44 (2008) 537–547

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cortex

Special issue: Original article

Switching between colors and shapes on the basis of positive and negative feedback: An fMRI and EEG study on feedback-based learning

Kiki Zanolie[a,b,1], *Santani Teng*[d,1], *Sarah E. Donohue*[e], *Anna C.K. van Duijvenvoorde*[a,b], *Guido P.H. Band*[a,b], *Serge A.R.B. Rombouts*[a,b,c] *and Eveline A. Crone*[a,b,*]

[a]*Leiden University Institute for Psychological Research (LU-IPR), Leiden University, Leiden, The Netherlands*
[b]*Leiden Institute for Brain and Cognition (LIBC), Leiden, The Netherlands*
[c]*Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands*
[d]*University of California, Davis, CA, USA*
[e]*Duke University, Durham, NC, USA*

ARTICLE INFO

Article history:
Received 19 March 2007
Reviewed 10 July 2007
Revised 30 August 2007
Accepted 17 November 2007
Published online 23 December 2007

Keywords:
Performance feedback
Hypothesis testing
Medial prefrontal cortex
Dorsolateral prefrontal cortex
fMRI
Intradimensional/extradimensional rule shift task

ABSTRACT

A crucial element of testing hypotheses about rules for behavior is the use of performance feedback. In this study, we used fMRI and EEG to test the role of medial prefrontal cortex (PFC) and dorsolateral (DL) PFC in hypothesis testing using a modified intradimensional/extradimensional rule shift task. Eighteen adults were asked to infer rules about color or shape on the basis of positive and negative feedback in sets of two trials. Half of the trials involved color-to-color or shape-to-shape trials (intradimensional switches; ID) and the other half involved color-to-shape or shape-to-color trials (extradimensional switches; ED). Participants performed the task in separate fMRI and EEG sessions. ED trials were associated with reduced accuracy relative to ID trials. In addition, accuracy was reduced and response latencies increased following negative relative to positive feedback. Negative feedback resulted in increased activation in medial PFC and DLPFC, but more so for ED than ID shifts. Reduced accuracy following negative feedback correlated with increased activation in DLPFC, and increased response latencies following negative feedback correlated with increased activation in medial PFC. Additionally, around 250 msec following negative performance feedback participants showed a feedback-related negative scalp potential, but this potential did not differ between ID and ED shifts. These results indicate that both medial PFC and DLPFC signal the need for performance adjustment, and both regions are sensitive to the increased demands of set shifting in hypothesis testing.

* *Corresponding author*. Department of Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.
E-mail address: ecrone@fsw.leidenuniv.nl (E.A. Crone).
[1] Shared first-authorship.

0010-9452/$ – see front matter © 2007 Elsevier Masson Srl. All rights reserved.
doi:10.1016/j.cortex.2007.11.005

1. Introduction

In order to adjust our behavior to changing circumstances in daily life, we must monitor the outcomes of our own actions. This type of performance monitoring is especially necessary when discriminating which behavior is appropriate and should be continued, and which behavior is inappropriate and should be adjusted. Therefore, we learn on the basis of *positive and negative feedback*. For example, feedback monitoring is important when we need to adjust pre-specified rules for behavior (Miltner et al., 1997) or when we need to test hypotheses about which behavior is currently appropriate (Barcelo and Knight, 2002).

Research into event-related potentials has demonstrated a differential neural response associated with receiving positive and negative performance feedback. For example, Miltner et al. (1997) asked participants to estimate a 1-sec time interval, which was followed by positive or negative feedback based on the accuracy of the estimate. They demonstrated that negative feedback was followed by a negative brain potential, which was observed approximately 230–270 msec following the presentation of negative feedback. This potential is maximal at frontocentral locations and has similarities with the error-related negativity (ERN), observed after a response error (Falkenstein et al., 1991). Several other studies have reported this brain potential in association with the presentation of negative feedback (Hajcak et al., 2006; Holroyd et al., 2006), which is thought to reflect an evaluation process that monitors expected and unexpected events (Nieuwenhuis et al., 2004). Therefore, this potential may reflect general performance monitoring and has been referred to as the feedback-ERN.

Source localization studies have suggested that the feedback-ERN originates in the medial prefrontal cortex (medial PFC), in or near the anterior cingulate cortex (ACC), and it has been suggested that the feedback-ERN reflects a dopaminergic learning signal (Holroyd and Coles, 2002; Miltner et al., 1997). Subsequent functional magnetic resonance imaging (fMRI) evidence, however, remains inconclusive about the role of the medial PFC in feedback-driven learning (Nieuwenhuis et al., 2005). Although some fMRI studies have shown increased activation in the caudal ACC following negative relative to positive feedback (Holroyd et al., 2004; Mars et al., 2005), others have failed to replicate this effect (Nieuwenhuis et al., 2005; van Veen et al., 2004). Thus, it is currently unclear how medial PFC, including ACC, is involved in the processing of positive and negative feedback when testing hypotheses.

Neuropsychological studies have emphasized the role of lateral PFC (lat-PFC) in feedback-driven learning, as demonstrated by perseverative behavior following feedback-induced rule switching in patients with damage to lat-PFC (Barcelo and Knight, 2002) and deviant ERN responses following damage to lat-PFC (Gehring and Knight, 2000). Kerns et al. (2004) argued that medial PFC signals response conflict and predicts activation in lat-PFC and associated performance adjustment. Therefore, in learning to adjust performance, medial PFC may be sensitive to the general information that signals that performance should be adjusted (negative *vs* positive feedback), whereas lat-PFC may be sensitive to the need to implement goal-directed and controlled behavior (Miller and Cohen, 2001). The goal of this study is to test the relative contributions of these brain regions to the processing of valence of performance feedback and the need for control.

One way to manipulate the demand for cognitive control is by the use of intradimensional (ID) versus extradimensional (ED) rule switches (O'Reilly et al., 2002). The classic ID/ED task (Dias et al., 1997; Roberts et al., 1988) is based on the Wisconsin Card Sorting Task (WCST) in that it demands a categorization of rules according to pre-specified rules. The task involves two kinds of switches: ID, in which the switch represents a category of the same dimension (e.g., switch from color to another color); and ED, in which a switch is made to a different dimension in the context of stimuli sharing the same general set of features (e.g., switch from color-to-shape). An ID switch therefore involves changing the target stimulus to another within the same dimension category, and an ED switch involves changing the target stimulus to one with a different dimension.

Dias et al. (1997) demonstrated that dorsolateral frontal lesions selectively impaired ED switches. This finding is consistent with prior studies which have suggested that lat-PFC is important for the transformation of higher levels of task rules into action, as in abstract rule switching (Cools et al., 2004). Hampshire and Owen (2006) reported that ventrolateral (VL) PFC, rather than DLPFC, played a central role in ED shifting. This result is consistent with prior reports which have suggested that VLPFC is important for the inhibition of the previously relevant response (Robbins, 2007). In contrast, they showed that the DLPFC is generally involved in solution search, a process in which the participant is actively searching the correct target response. This finding concurs with previous models of DLPFC function in which a role for the DLPFC in functions such as monitoring within working memory (e.g., Petrides, 2000) was suggested. In this study we were specifically interested in the feedback processing after having to switch to a new rule dimension (i.e., an ED condition) relative to the same rule dimension (i.e., ID condition). Based on these previous findings (Cools et al., 2004; Dias et al., 1997; Hampshire and Owen, 2006), we predict that lat-PFC will be especially sensitive to feedback-based learning following a switch to ED rules relative to ID rules.

In this study, we examined the role of medial PFC and lat-PFC in relation to feedback-based rule learning using a rule learning task that was inspired by the ID/ED switch task. The same participants were asked to participate in two sessions; fMRI data were assessed on the first occasion and EEG recordings on the second. In both sessions, a predictable switch task was performed which consisted of pairs of trials: a switch trial followed by a repetition trial. Prior to each trial pair, participants were cued to sort two nameable images on the basis of color or shape (see Fig. 1). The response was followed by a positive feedback signal (+) or a negative feedback signal (x), and participants were instructed to use this information to make the correct choice on the repetition trial. As the task required an attentional switch rather than a switch of response mappings, each trial pair presented two new stimuli in two different colors to control for the possible confound of response interference. ED conditions are more likely to require greater monitoring because it is necessary to overcome interference from the previous dimension. However, this modification made the task different from the original ID/ED switch task, because

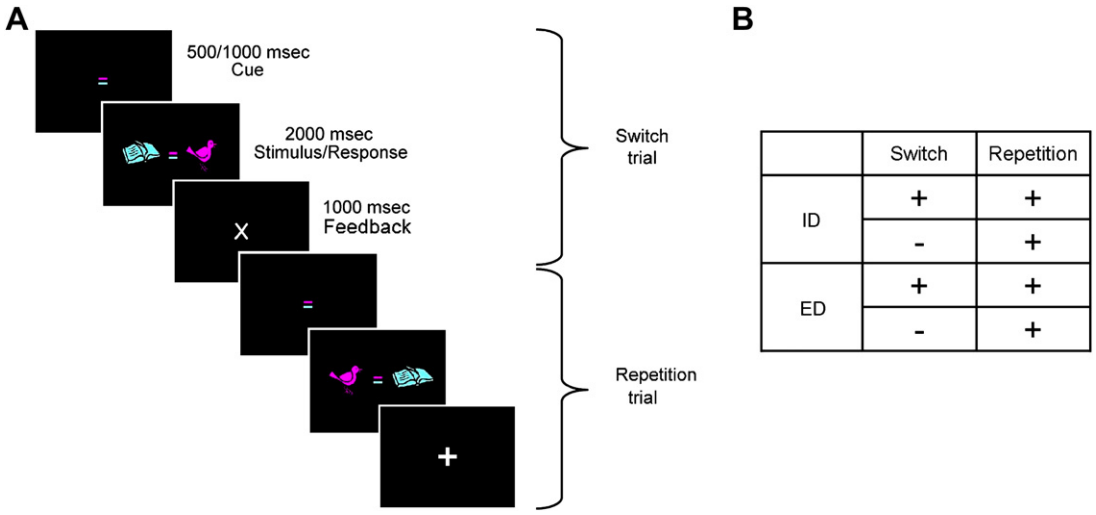

Fig. 1 – Task design and experimental conditions. (A) Trials were paired into switch and repetition trials. A cue indicated the dimension to select (shape or color) and was followed by the stimulus presentation. A response was followed by positive or negative feedback, after which the sequence repeated itself during the repetition trial. Trials were separated by intertrial intervals (not shown) of 2000–8000 msec. (B) Four switch conditions were analyzed: ID shift followed by positive feedback, ID shift followed by negative feedback, ED shift followed by positive feedback, ED shift followed by negative feedback. Switch trials that were followed by incorrect repetition trials were excluded from analysis.

there was no *response* interference from the previous task set. We chose to use this manipulation to focus specifically on feedback-based learning rather than stimulus or response reversal learning. Therefore, ID and ED conditions were mainly used to manipulate difficulty/monitoring level.

The current design allowed us to examine neural and electrophysiological responses to positive and negative feedback when this information was used to test hypotheses *following* an ID or an ED switch. Therefore, the analyses focused on the feedback events following the first switch trial of each trial pair. We predicted that medial PFC would be sensitive to the valence of the feedback (positive or negative) (Holroyd et al., 2004), and lat-PFC was expected to be sensitive to the dimension of the switch (ID or ED) (O'Reilly et al., 2002). The subsequent recording of the feedback-ERN allowed us to test whether this brain potential is modulated by the dimension of the switch, or whether is it sensitive to negative feedback per se.

2. Methods

2.1. Subjects

Eighteen healthy right-handed paid volunteers (11 females, 7 males; age 18–24, mean = 22.3) participated in the fMRI experiment. Of these volunteers, 15 (10 females, 5 males; age 20–24; mean = 22.6) participated in the EEG experiment. One of the participants in the EEG session was excluded due to a high level of noise in the data. Subsequent EEG analyses consist of the remaining 14 participants. All participants reported normal or corrected-to-normal vision and the absence of neurological or psychiatric impairments. For MRI safety reasons all participants were screened for any metal in or on their bodies. All participants gave informed consent for the study, and all procedures were approved by the Leiden University, Department of Psychology and the medical ethical committee of the Leiden University Medical Center. Standard intelligence scores were obtained from each participant using the Raven's Progressive Matrices test. All participants had average or above average IQ scores.

2.2. Task

In both the fMRI and EEG sessions, participants performed an Intradimensional/Extradimensional (ID/ED) Rule Task in which they used positive and negative feedback to determine and apply a simple rule for target stimulus selection. Stimuli consisted of four pairs of nameable pictures whose positions (left *vs* right side of a central fixation cross) and colors were interchangeable within each pair (Fig. 1). A cue alerted the participant to the selection criterion, either the shape (white equals sign) or color (colored equals sign) of the stimulus. Stimuli were presented in pairs of trials. In the first trial (referred to below as the switch trial), participants have been cued to the selection criterion but did not know the rule designating the target. For example, a participant cued to choose a stimulus based on color might randomly pick the blue item. Participants' responses generated positive or negative feedback and were followed by the second trial (referred to below as the repetition trial), in which they were again cued to the selection criterion before seeing the paired stimuli (the same items, with the colors and positions switched 50% of the time). The participants then chose the correct item based on feedback they had received in the previous trial. Again, these choices were followed by positive or negative feedback. A trial was classified as ID when it featured the same selection criterion as the previous pair of trials (i.e., color-to-color or shape-to-shape).

An ED trial occurred when the selection criterion changed. Participants were instructed to respond as accurately as possible.

Thus, in contrast to the classic ID/ED Switch Task (Dias et al., 1997) the switches were not triggered by negative feedback following a change in reward contingencies, but were predictable for every other trial. Furthermore, the negative feedback following the switch trial did not indicate any real 'error' in behavior by participants but merely that they had properly guessed at known chance between two alternatives. Finally, the participants could correct their choice with near 100% accuracy on the next trial.

The combined stimulus presentation and response window was fixed at 2000 msec. The pre-stimulus cue period was 500 msec in the fMRI experiment and 1000 msec in the EEG experiment. The button press selecting a target generated 1000 msec of either positive or negative feedback. Intertrial intervals were jittered exponentially varying from 2000, 4000, 6000 and 8000 msec in approximately 25% of the trials, where 2000 msec is the most and 8000 msec the least prevalent (Dale, 1999). In all cases, participants were informed that their task in the experiment would be to make some simple decisions about pairs of pictures. Each participant underwent a behavioral practice session to ensure proficiency in the task prior to the respective fMRI and EEG sessions. In the fMRI session participants completed four experimental blocks, while in the EEG session five blocks were completed, all consisting of 100 trials each.

2.3. MRI procedure

fMRI data were acquired with a standard whole-head coil on a 3.0 T Philips Achieva MRI scanner (Eindhoven, Netherlands) at the Leiden University Medical Center. Localizer and T2 structural scans were obtained for each participant. Stimuli were projected onto a screen located at the head of the scanner bore and viewed by participants by means of a mirror mounted to the head coil assembly. T2*-weighted echoplanar images were obtained during four functional runs of 232 volumes each, of which the first two were discarded to allow for equilibration of T1 saturation effects. Each volume covered the whole brain (38 slices of thickness 2.75 mm, field of view 220 mm, 80×80 matrix, in-plane resolution 2.75 mm) and was acquired every 2211 msec (TE = 30 msec, ascending interleaved acquisition). A high-resolution T1-weighted anatomical scan was obtained from each participant after the functional runs. In accordance with Leiden University Medical Center policy, all anatomical scans were reviewed and cleared by the radiology department following each scan. No anomalous findings were reported.

2.4. EEG procedure

All electrodes were pre-amplified Ag/AgCl BioSemi® electrodes. DC recorded time series were digitized in 24-bit format with a resolution of 31nV at 256 Hz. EEG was recorded from 18 electrodes; AF3, AFz, AF4, F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4, P3, Pz, P4; embedded in a head-cap over the scalp according to the 10–20 system, and the two mastoids. Eye-movements and blinks were monitored by bipolar EOG recordings of the left versus right outer canthus (horizontal) and left supraorbital versus infraorbital (vertical).

2.5. EEG data analysis

EEG signals were referenced offline to the average mastoid. First, EEG signal was high-pass filtered at 0.1 Hz (24 dB/oct) to remove low-frequency activity that could cause serious drift, harming EOG correction procedures. After filtering, EEG signal was corrected for ocular artifacts with the Gratton et al. (1983) algorithm. EEG signal was low-pass filtered at 16 Hz (24 dB/oct) before segmentation. Segments of 1000 msec of data (200 msec baseline) were extracted separately for ID switch trials where participants received negative and ID switch trials where participants received positive feedback. The same segments were extracted for ED switch trial types, resulting in the four specified conditions. All segments were synchronized to the onset of feedback presentation.

The feedback-ERN was defined as the average amplitude of the waveform in a window from 220–300 msec after feedback presentation relative to its immediate preceding positivity between 130 and 200 msec. These time windows were chosen based on previous literature for feedback-ERN components (e.g., Gehring and Willoughby, 2002; Hajcak et al., 2006; Holroyd and Coles, 2002; Nieuwenhuis et al., 2005) and supplementary visual inspection of grand-averaged waveforms. Difference scores were used in a separate repeated measures ANOVA on locations of interest (Fz and FCz) for the within-subject factors Feedback (negative *vs* positive) and Type of Switch (ID *vs* ED).

2.6. fMRI data analysis

Data were preprocessed using SPM2 (Wellcome Department of Cognitive Neurology, London). Images were corrected for differences in timing of slice acquisition, followed by rigid-body motion correction. Functional volumes were spatially normalized to EPI templates. The normalization algorithm used a 12-parameter affine transformation together with a nonlinear transformation involving cosine basis functions and resampled the volumes to 3 mm cubic voxels. Templates were based on the MNI305 stereotaxic space (Cocosco et al., 1997), an approximation of Talairach space (Talairach and Tourneaux, 1988). Functional volumes were spatially smoothed with an 8 mm FWHM isotropic Gaussian kernel. Statistical analyses were performed on individual subjects' data using the general linear model in SPM2. The fMRI time series data were modeled by a series of events convolved with a canonical hemodynamic response function (HRF). The feedback stimulus of each trial (switch and repetition trials) was modeled as a zero-duration event. The trial functions were used as covariates in a general linear model, along with a basic set of cosine functions that high-pass filtered the data, and a covariate for session effects. The least-squares parameter estimates of height of the best-fitting canonical HRF for each condition were used in pairwise contrasts. The resulting contrast images, computed on a subject-by-subject basis, were submitted to group analyses. At the group level, contrasts between conditions were computed by performing one-tailed *t*-tests on these images, treating subjects as a random effect. Task-related responses were considered significant if they consisted of at least 10 contiguous voxels that

exceeded an uncorrected threshold of $p < .001$. We used an uncorrected threshold to explore possible differences in activation at a liberal threshold, but used regions of interest analysis to test for region × condition effects. Region-of-interest (ROI) analyses were performed to further characterize rule sensitivity of four a priori predicted regions (left and right DLPFC and left and right medial PFC). ROI analyses were performed with the Marsbar toolbox in SPM2 (Brett et al., 2002, http://marsbar.sourceforge.net). ROIs that spanned several functional brain regions were subdivided by sequentially masking the functional ROI with each of several anatomical Marsbar ROIs. We made use of anatomical template ROIs included with the marsbar program (http://marsbar.sourceforge.net). The contrast used to generate functional ROIs was based on the F-contrast negative feedback versus positive feedback across dimensions, $p < .05$, corrected for multiple comparisons. For all ROI analyses, effects were considered significant at an α of .05. For each ROI, the center of mass is reported.

3. Results

3.1. Behavioral results

As anticipated, participants' performance did not significantly differ from chance level on the switch trial in either the fMRI or the EEG sessions. The feedback was successfully used to find the appropriate rule, as indicated by high accuracy on the repetition trial in each respective pair. In the fMRI session, accuracy and RTs on repetition trials were compared for trials that followed an ID versus ED switch, and trials that followed negative versus positive feedback (see Fig. 2). The ANOVA showed that accuracy was slightly but significantly higher for ID switch rules relative to ED switch rules ($F(1,17) = 11.88$, $p < .005$), as well as positive relative to negative feedback trials ($F(1,17) = 14.68$, $p < .001$), but there was no dimension × feedback interaction ($p > .92$). A similar analysis for reaction times (RTs) showed that RTs were faster following positive feedback than following negative feedback, ($F(1,17) = 16.55$, $p < .001$), but there were no effects of dimension ($p > .45$).

These differences were not significant in the EEG session for accuracy or response latencies ($F < 1$).

3.2. fMRI results: whole-brain comparisons

The whole-brain comparisons focused on the difference in brain activation following negative relative to positive feedback on the switch trial, for both ID and ED switches. The comparison negative > positive feedback for ID switches resulted in a cluster of activation in medial PFC. As expected, this activation was also present in the comparison negative > positive feedback for ED switches, but the ED comparison additionally resulted in activation in right DLPFC (Table 1, Fig. 3). We performed ROI analyses to examine these differences in more detail, as discussed below.

3.3. fMRI results: ROI comparison

At first glance, it appears that negative feedback results in increased activation in medial PFC and lateral PFC, but that this effect was larger following ED relative to ID switches. To examine the relative contributions of these areas, an ROI analysis was performed for left and right medial PFC and left and right DLPFC.

We examined the effects of positive and negative feedback on the switch trial following both ID and ED switches with a Feedback (2) × Dimension (2) × Region (2) × Hemisphere (2) ANOVA (see Fig. 4). This ANOVA showed that both regions were more active following negative than positive feedback ($F(1,17) = 22.87$, $p < .001$) and revealed a significant Feedback × Dimension interaction ($F(1,17) = 6.59$, $p < .01$). The latter effect indicates that across regions the difference between activation following negative relative to positive feedback was larger for ED trials (difference = .83) than following ID trials (difference = .33) (see Fig. 4). Contrary to expectations, there were no interactions with Region (all $p > .52$).

A closer inspection of the time courses, however, demonstrated that the difference between the peak of negative feedback-related activation occurred earlier for medial PFC than for DLPFC. For this exploratory analysis, we performed a Region (medial PFC *vs* DLPFC) × Hemisphere (left *vs*

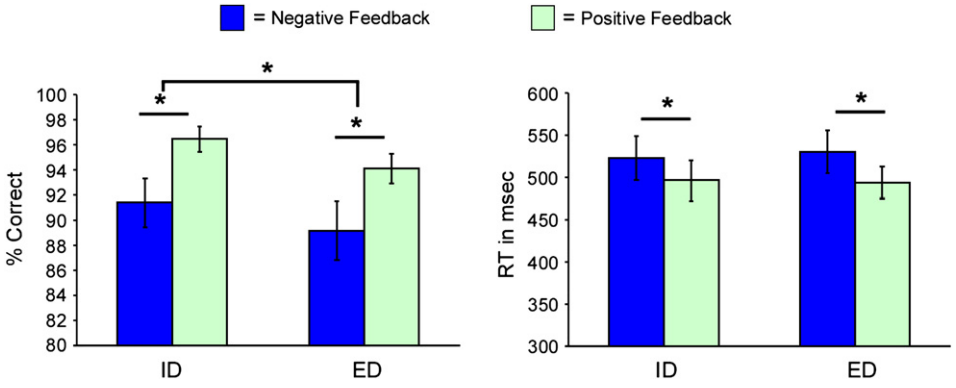

Fig. 2 – Behavioral results for fMRI session for repetition trials. Participants were significantly more accurate for the ID trials compared to the ED trials, and on trials following positive feedback relative to negative feedback. RT analysis demonstrated that participants slowed responses following negative relative to positive feedback.

Table 1 – Coordinates for comparison of extradimensional negative feedback > extradimensional positive feedback, and intradimensional negative feedback > intradimensional positive feedback, at $p < .001$ (uncorrected), min 10 active voxels

	BA	x	y	z	Z-value
Extradimensional negative FB > positive FB					
R IFG	45	48	30	6	4.60
R MFG	46	51	30	24	4.47
R MFG	46	48	42	21	3.59
SFG	6	0	15	63	4.50
L SFG	8	−6	15	51	4.28
L SFG	6	−9	6	57	4.14
R Sup Parietal	7	27	−66	48	3.99
R Inf Parietal	40	39	−45	42	3.84
R Precuneus	7	12	−60	51	3.83
L Precuneus	7	−9	−78	48	3.97
L Precuneus	7	−12	−75	57	3.85
L Precuneus	7	−15	−66	48	3.97
L Sup Parietal	7	−33	−66	48	3.16
L Sup Parietal	7	−39	−57	48	3.16
L Inf Parietal	40	−42	−45	45	3.21
R MFG	6	30	12	54	3.45
L Insula	13	−36	15	6	3.45
Intradimensional negative FB > positive FB					
L MFG	6	−42	3	39	4.21
L MTG	21	−51	−33	−6	3.73
L SFG	6	−9	−3	60	3.62
L SFG	6	−3	6	60	3.59

IFG = inferior frontal gyrus; MFG = middle frontal gyrus; SFG = superior frontal gyrus; MTG = middle temporal gyrus.

right) × Dimension (ID *vs* ED) × Feedback (positive *vs* negative) × Time Point (eight time points) repeated measures ANOVA, which resulted in a significant Region × Time Point interaction ($F(7,119) = 10.79$, $p < .001$). There were no interactions with dimension or hemisphere, but the Region × Feedback × Time Point interaction approached significance ($F(7,119) = 1.94$, $p = .07$). As can be seen in Fig. 5, medial PFC showed a larger and earlier peak in activation than DLPFC following negative feedback (Region × Time Point interaction, $F(7,119) = 9.58$, $p < .001$). Post hoc comparisons for each time point confirmed a larger difference between medial PFC relative to DLPFC for time point 6.6 sec (main effect region, $F(1,17) = 22.61$, $p < .001$). There was also a difference in the pattern of activation of medial PFC versus DLPFC following positive feedback (Region × Time Point interaction, $F(7,119) = 5.93$, $p < .001$), but these values showed a more variable pattern.

3.4. *Brain-behavior correlations*

To examine the relation between performance and brain activation, we examined the correlations between the differences scores for accuracy (% correct following positive feedback−% correct following negative feedback), RTs (RTs following negative feedback−RTs following positive feedback) and brain activation (ROI value negative feedback trials−ROI value positive feedback trials) for each ROI. These correlations demonstrated that activation in DLPFC was associated with a larger difference in accuracy, for both the left ($r = .62$, $p < .01$) and right ($r = .52$, $p < .05$) DLPFC. There were no significant correlations with RT differences. In contrast, activation in medial PFC was associated with increased difference in response latency, for both left ($r = .52$, $p < .05$) and right ($r = .53$, $p < .05$) medial PFC. There were no significant correlations with accuracy.

3.5. *EEG results*

Similar Dimension (ID *vs* ID) × Feedback (positive *vs* negative) ANOVAs were performed for the peak of activation in the 220–300 msec time window after feedback presentation, relative to the preceding 130–200 msec time window in the EEG data. In Fig. 6, the feedback-related data are plotted separately for each dimension switch. We focused on two locations of interest, Fz and FCz. The ANOVA for Dimension (2) × Feedback (2) for the peak-to-peak difference scores did not result in a significant interaction. However, at FCz negative feedback did result in a significantly larger deflection than positive feedback conditions ($F(1,13) = 6.59$, $p < 0.05$). Correlations between EEG deflections and ROI activation in medial PFC and DLPFC were not significant.

4. Discussion

The present results demonstrate recruitment of medial PFC and lat-PFC in feedback processing. Both medial PFC and lat-PFC were sensitive to negative feedback, and more so when this feedback followed an ED shift, relative to an ID shift. Both regions correlated with subsequent performance adjustment, strengthening our hypothesis that medial PFC and lat-PFC are important for feedback-based learning. Detailed ROI analyses demonstrated that, contrary to expectations, medial PFC and lat-PFC did not differ in sensitivity to negative and positive performance feedback following ED and ID shifts, but did differ in timing of peak activation.

The data concur with earlier studies showing that medial PFC is recruited following the presentation of negative performance feedback (Holroyd et al., 2004; Mars et al., 2005), and with studies suggesting that medial PFC is sensitive to the first prediction that performance should be adjusted (Holroyd and Coles, 2002). The center of activation was in SMA rather than ACC. However, prior studies on feedback learning have also reported activation that was in (pre-)SMA rather than ACC (Holroyd et al., 2004, coordinates $x = 4$, $y = 12$, $z = 59$; in this study, $x = -4$, $y = 11$, $z = 57$). It should be noted that ACC and (pre-)SMA may have separable roles in feedback processing. For example, a study by Mars et al. (2005) showed that the anterior rostral cingulate zone (RCZa, coordinates $x = 8$, $y = 30$, $z = 32$), a specific portion of the medial PFC, responds to both internal and external error sources and is activated in response to the first signal that an error has been made. This evaluative information can be used to adapt behavior accordingly (Holroyd et al., 2004). In contrast, the pre-SMA region (coordinates $x = 8$, $y = 10$, $z = 55$) showed response-related effects over learning-dependent modulations of activity in both correct and incorrect trials, suggesting that pre-SMA might be involved in a motor aspect of a learning

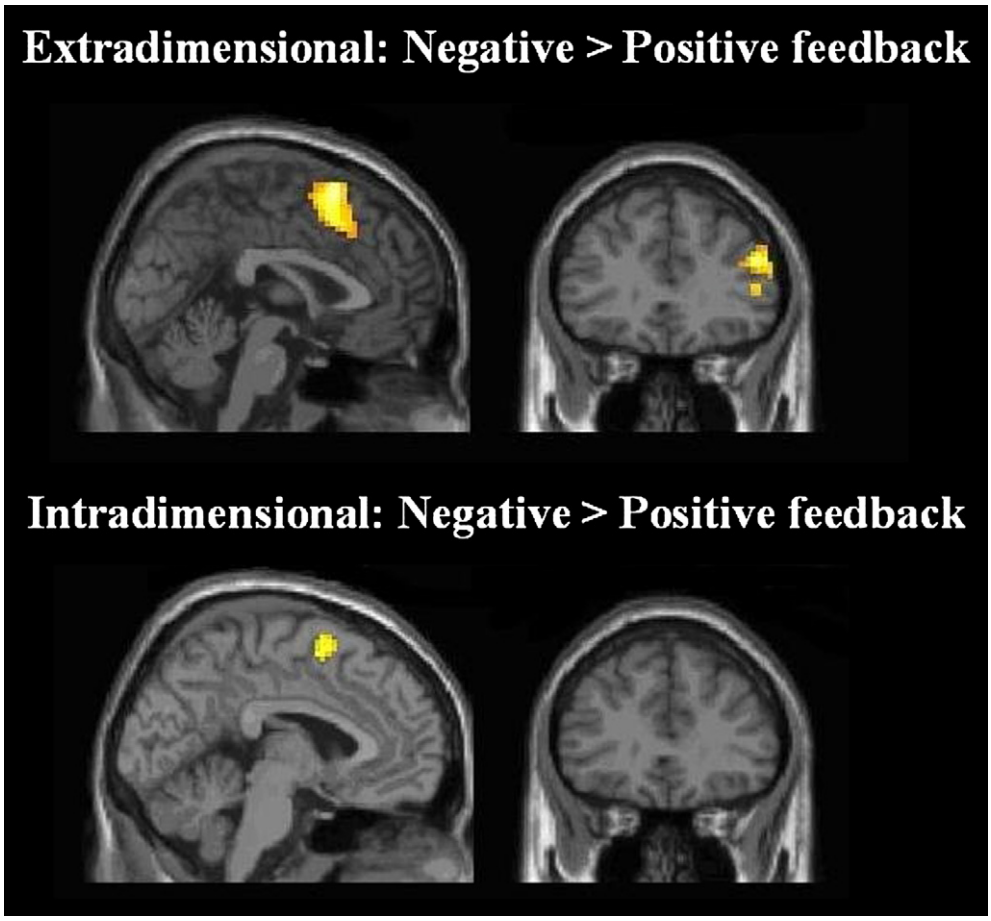

Fig. 3 – The top panel shows the contrast for negative > positive feedback for the ED switch condition at X = −2, Y = 32, Z = 33. Clusters of activation shown include the medial frontal gyrus (BA 6; X = 0, Y = 15, Z = 63) (left) and right DLPFC (BA 46; X = 51, Y = 30, Z = 24). The bottom panel depicts the same contrast for the ID switch trials. Here, activation was only observed in the left medial frontal gyrus (BA 6; X = −3, Y = 6, Z = 60). Other clusters of significant activation found in these contrasts (not shown) are listed in Table 1.

process. In this study we could not dissociate valence from informative value of the feedback. In future studies, it will be important to specify the role of subregions within the medial wall in greater detail (see also Zanolie et al., 2007).

The current study differs from prior studies that failed to report medial PFC activation following negative feedback in that these studies demanded a single response (time estimation) (Nieuwenhuis et al., 2005; van Veen et al., 2004), whereas in the current study a response selection was necessary. It was previously shown that medial PFC is sensitive to situations in which the number of response alternatives is large (Walton et al., 2004; Zanolie et al., 2007). Therefore, negative feedback may signal the need to make a response selection under uncertain conditions.

Besides medial PFC, lat-PFC was also sensitive to negative feedback information, consistent with prior reports showing that damage to lat-PFC results in impaired feedback processing (Gehring and Knight, 2000), or response adjustment (Barcelo and Knight, 2002). Moreover, in the current study we manipulated the demands for complexity of feedback processing by signaling the need for ED versus ID shifts. The whole-brain comparisons demonstrated that, similar to prior reports in animals (Dias et al., 1997) and fMRI studies in humans (Cools et al., 2002; Hampshire and Owen, 2006), negative feedback following an ED shift cue resulted in more activation in lat-PFC than did negative feedback associated with an ID shift cue. This difference was also observed for medial PFC, suggesting that both regions are important for feedback processing that follows dimension shifting. In the current study, we compared feedback processing following ED attention switches to ID attention switches. Prior studies that have compared ED shifting relative to reversal shifts indicated that these processes are qualitatively different and implicate a dissociation between the relative contribution of dorsolateral and ventral PFC, respectively (e.g., Cools et al., 2002; Dias et al., 1996; Robbins, 2007). It should be noted that this study did not examine the shifting process, but mainly focused on feedback processing after more or less heavy demands on attention shifting (as manipulated by ED and ID conditions). The positive correlation between activation levels in medial PFC/lat-PFC ROIs and response cost corroborate our observation that these regions were sensitive to the difficulty

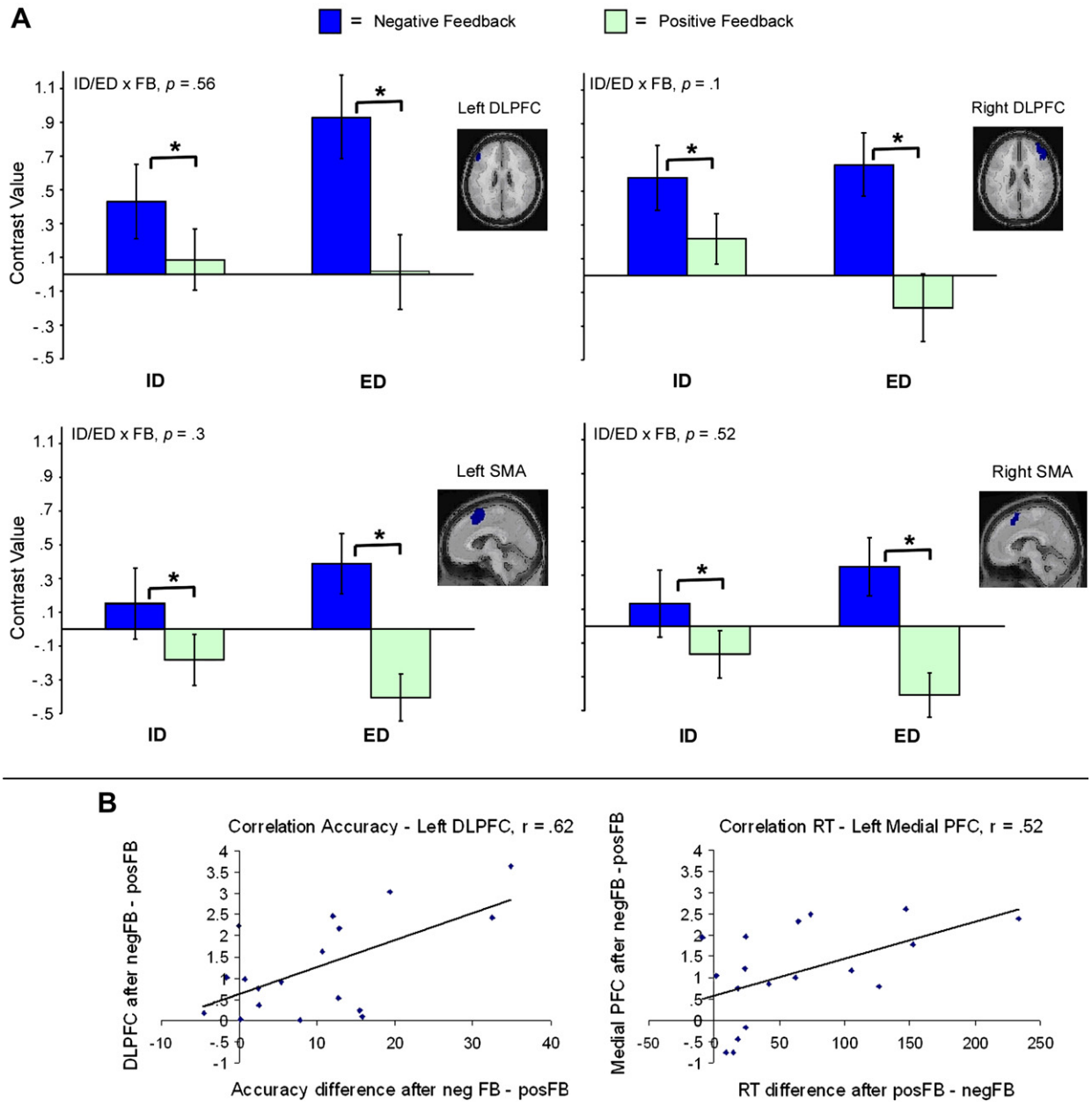

Fig. 4 – (A) ROI analyses. Top: activation pattern for DLPFC (left DLPFC, BA 9: $x = -48$, $y = 33$, $z = 33$; right DLPFC, BA 9/46: $x = 51$, $y = 36$, $z = 24$). Contrast values for the negative and positive feedback conditions for the ID and ED switch trials. In both trial types, the activity was significantly greater for the negative than positive feedback. Bottom: activation pattern for medial PFC (left SMA, BA 6: $x = -4$, $y = 11$, $z = 57$; right SMA, BA 6: $x = 7$, $y = 14$, $z = 59$). As above, this region was significantly more engaged for negative than positive feedback across both conditions. (B) Correlations between accuracy cost and activation in left DLPFC ROI, and RT cost and activation in left medial PFC.

demands rather than implementation of control. This specific focus (on feedback processing rather than shifting) can explain why we failed to find a dissociation between medial PFC and lat-PFC. In future studies, it will be of interest to examine the interplay between medial PFC and lateral PFC in the context of ED, ID and reversal shifts (see also Hampshire and Owen, 2006).

Although we failed to dissociate between the roles of medial PFC and lat-PFC in feedback processing with our paradigm, we did find potentially interesting differences in the timing of brain activation between the regions, as well as in the type of brain-behavior correlations. Time series comparisons indicated that activation associated with negative feedback peaked earlier in medial PFC than in lat-PFC. This result is consistent with the hypothesis that medial PFC signals the need for cognitive control, whereas lat-PFC is important for producing behavioral adjustments (Holroyd and Coles, 2002; Kerns et al., 2004; Miller and Cohen, 2001). In a prior study by Kerns et al. (2004), the dependency of medial PFC and lat-PFC was demonstrated using a Stroop

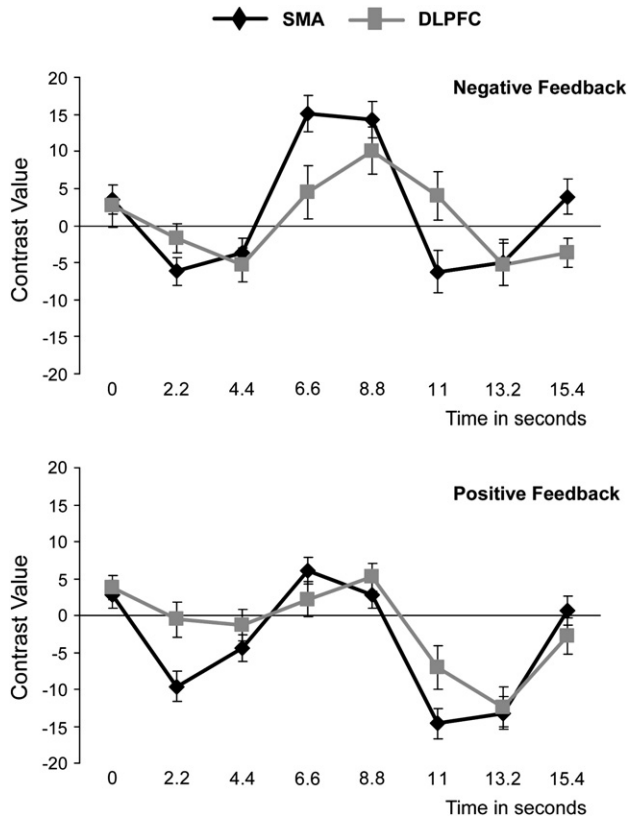

Fig. 5 – The graph shows a plot of the time course of activation for medial PFC (SMA) and DLPFC following negative and positive feedback (ROIs from Fig. 4, values averaged across left and right hemispheres).

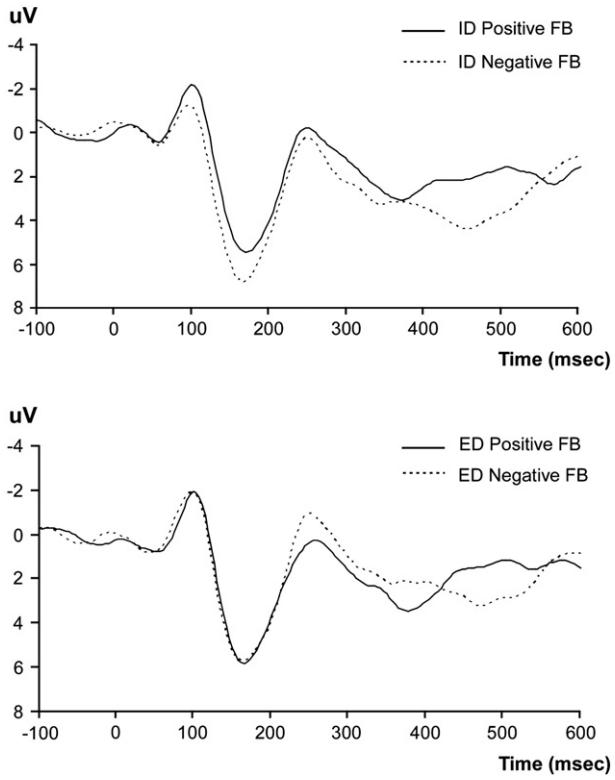

Fig. 6 – ERP results. Analysis was time locked to the onset of the feedback. Averaged ERPs are plotted for site Fz (top) and FCz (bottom) for ED (left) and ID (right) trials. Waveforms are shown for positive and negative feedback on switch trials. At Fz, for ED switches, the second negative deflection was significantly larger for negative feedback compared to positive feedback conditions on switch trials. At FCz, the second negative deflection was significantly larger for negative feedback compared to positive feedback conditions for both ID and ED switch trials.

task, in which increased activation in medial PFC (particularly ACC) resulted in increased lat-PFC activation and performance adjustments on subsequent trials. In this study, we show that this account is also plausible for feedback processing. However, these findings should be interpreted with caution, given that regions may differ in BOLD timing independently of the demands on cognitive control (Handwerker et al., 2004). Further, although both brain regions correlated positively with response cost, DLPFC correlated with accuracy and medial PFC with response times on subsequent trials. These correlations indicate that lat-PFC and medial PFC signaled the difficulty level of feedback processing (as indicated by a positive correlation with response cost) rather than the implementation of control (in which case we could expect a negative correlation). The positive correlations can be explained by greater proactive interference from previous trials which result in increased brain activation and decreased accuracy. In future studies, it will be interesting to test the functional connectivity between medial PFC and DLPFC in relation to rule learning and hypothesis testing.

Analysis of time-specific ERP components demonstrated that the feedback-ERN that followed negative feedback did not differentiate between ID and ED shifts. The amplitude of the feedback-ERN did not correlate with activation in either medial PFC or lat-PFC, which is consistent with prior studies suggesting that the feedback-ERN has a source outside medial PFC (van Veen et al., 2004). Although the feedback-ERN was observed in a time window similar to previous studies (Hajcak et al., 2006; Holroyd et al., 2006), its amplitude was smaller compared to previous reports (for a review, see Nieuwenhuis et al., 2004). This difference can be associated with several differences in the experimental design. First, in this study the negative feedback did not have motivational significance (correct *vs* incorrect), but was presented with a 50% probability after a switch trial, and both positive and negative feedback were informative. Prior research suggested that the feedback-ERN is sensitive to a binary representation of good versus bad outcomes, not to the informative value of this representation (Hajcak et al., 2006). Therefore, it is possible that the feedback-ERN is less sensitive to information related to hypothesis testing. Second, prior studies have suggested that the feedback-ERN is associated with a negative prediction error signal (Holroyd and Coles, 2002). In this study, negative feedback after a switch trial was presented with a 50% probability, similar to the study reported by Miltner et al. (1997). However, contrary to Miltner et al., participants had no a priori prediction about the success of their performance,

suggesting that the sensitivity of the feedback-ERN may be dependent on the extent to which participants have predictions about the outcome of their choice. Finally, all participants performed the task in the fMRI session before returning for the EEG session. Therefore, it is possible that effects are confounded by practice effects. In future studies, it is important to counterbalance the fMRI and EEG sessions.

Together, the current results support the hypothesis that both medial PFC and lat-PFC are important for feedback processing. Both regions were sensitive to the valence of the feedback (negative *vs* positive) and both regions were sensitive to the difficulty of the prior switch (ED shifts *vs* ID shifts). Interestingly, there seems to be a difference in the timing of activation, with medial PFC peaking earlier than DLPFC. Additionally, both regions correlated with performance, but lat-PFC correlated with subsequent accuracy, whereas medial PFC correlated with subsequent slowing. These findings hint that medial PFC and lat-PFC may be differentially important for performance adjustment. The feedback-ERN was larger for negative than positive feedback, but did not differentiate between ID and ED shifts, suggesting that this component is sensitive to valence only. In our prior developmental studies, we demonstrated that the ability to detect the valence of feedback matures earlier than the ability to adjust performance based on feedback (Crone et al., 2004, 2006). This finding, together with the current findings, has led to the hypothesis that these two processes may be supported by different brain regions, which may have separate developmental trajectories. These hypotheses are currently being tested in our laboratory (van Duijvenvoorde et al., submitted for publication).

REFERENCES

Barcelo F and Knight RT. Both random and perseverative errors underlie wcst deficits in prefrontal patients. *Neuropsychologia*, 40: 349–356, 2002.

Brett M, Anton JL, Valabregue R and Poline JB. Region of interest analysis using an spm toolbox. In: Paper presented at the eighth international conference on functional mapping of the human brain. Sendai, Japan; 2002.

Cocosco CA, Kollokian V, Kwan RKS, and Evans AC. Brain web: online interface to a 3D MRI simulated brain database. *Neuroimage*, 5: S452, 1997.

Cools R, Clark L, Owen AM, and Robbins TW. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience*, 22: 4563–4567, 2002.

Cools R, Clark L, and Robbins TW. Differential responses in human striatum and prefrontal cortex to changes in object and rule relevance. *The Journal of Neuroscience*, 24: 1129–1135, 2004.

Crone EA, Jennings JR, and van der Molen MW. Developmental change in feedback processing as reflected by phasic heart rate changes. *Developmental Psychology*, 40: 1228–1238, 2004.

Crone EA, Somsen RJ, Zanolie K, and van der Molen MW. A heart rate analysis of developmental change in feedback processing and rule shifting from childhood to early adulthood. *Journal of Experimental Child Psychology*, 95: 99–116, 2006.

Dale AM. Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8: 109–114, 1999.

Dias R, Robbins TW, and Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, 380: 69–72, 1996.

Dias R, Robbins TW, and Roberts AC. Dissociable forms of inhibitory control within prefrontal cortex with an analog of the wisconsin card sort test: restriction to novel situations and independence from ''on-line'' processing. *Journal of Neuroscience*, 17: 9285–9297, 1997.

Falkenstein M, Hohnsbein J, Hoormann J, and Blanke L. Effects of crossmodal divided attention on late erp components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, 78: 447–455, 1991.

Gehring WJ and Knight RT. Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, 3: 516–520, 2000.

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

Gratton G, Coles MGH, and Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55: 468–484, 1983.

Hajcak G, Moser JS, Holroyd CB, and Simons RF. The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology*, 71: 148–154, 2006.

Hampshire A and Owen AM. Fractionating attentional control using event-related fMRI. *Cerebral Cortex*, 16: 1679–1689, 2006.

Handwerker DA, Ollinger JM, and D'Esposito M. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21: 1639–1651, 2004.

Holroyd CB and Coles MG. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109: 679–709, 2002.

Holroyd CB, Hajcak G, and Larsen JT. The good, the bad and the neutral: electrophysiological responses to feedback stimuli. *Brain Research*, 1105: 93–101, 2006.

Holroyd CB, Nieuwenhuis S, Yeung N, Nystrom L, Mars RB, Coles MG, and Cohen JD. Dorsal anterior cingulate cortex shows fmri response to internal and external error signals. *Nature Neuroscience*, 7: 497–498, 2004.

Kerns JG, Cohen JD, Macdonald 3rd AW, Cho RY, Stenger VA, and Carter CS. Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303: 1023–1026, 2004.

Mars RB, Coles MG, Grol MJ, Holroyd CB, Nieuwenhuis S, Hulstijn W, and Toni I. Neural dynamics of error processing in medial frontal cortex. *Neuroimage*, 28: 1007–1013, 2005.

Miller EK and Cohen JD. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24: 167–202, 2001.

Miltner WHR, Braun CH, and Coles MG. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a ''generic'' neural system for error detection. *Journal of Cognitive Neuroscience*, 9: 788–798, 1997.

Nieuwenhuis S, Holroyd CB, Mol N, and Coles MG. Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience and Biobehavioral Reviews*, 28: 441–448, 2004.

Nieuwenhuis S, Slagter HA, Von Geusau NJ, Heslenfeld DJ, and Holroyd CB. Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience*, 21: 3161–3168, 2005.

O'Reilly RC, Noelle DC, Braver TS, and Cohen JD. Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cerebral Cortex*, 12: 246–257, 2002.

Petrides M. The role of the mid-dorsolateral prefrontal cortex in working memory. *Experimental Brain Research*, 133: 44–54, 2000.

Robbins TW. Shifting and stopping: fronto-striatal substrates, neurochemical modulation and clinical implications. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362: 917–932, 2007.

Roberts AC, Robbins TW, and Everitt BJ. The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *The Quarterly Journal of Experimental Psychology*, 40: 321–341, 1988.

Talairach J and Tourneaux P. *Co-Planar Stereotactic Atlas of the Human Brain*. New York: Thieme, 1988.

van Duijvenvoorde A, Zanolie K, Rombouts SARB, Raijmakers M and Crone EA. Valuing the positive or evaluating the negative? Bringing neuroscience in the classroom, submitted for publication.

van Veen V, Holroyd CB, Cohen JD, Stenger VA, and Carter CS. Errors without conflict: implications for performance monitoring theories of anterior cingulate cortex. *Brain and Cognition*, 56: 267–276, 2004.

Walton ME, Devlin JT, and Rushworth MF. Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience*, 7: 1259–1265, 2004.

Zanolie K, Leijenhorst L, Rombouts SARB, and Crone EA. Separable neural mechanisms contribute to feedback processing in a rule-learning task. *Neuropsychologia*, 2007.