European Journal of Neuroscience, Vol. 21, pp. 3161–3168, 2005

Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes

Sander Nieuwenhuis,[1] Heleen A. Slagter,[2] Niels J. Alting von Geusau,[1] Dirk J. Heslenfeld[1] and Clay B. Holroyd[3]
[1]Department of Cognitive Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands
[2]Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands
[3]Department of Psychology, University of Victoria, Victoria, Canada

Keywords: cingulate cortex, ERP, feedback, fMRI, human

Abstract

Previous research has identified a component of the event-related brain potential (ERP), the feedback-related negativity, that is elicited by feedback stimuli associated with unfavourable outcomes. In the present research we used event-related functional magnetic resonance imaging (fMRI) and electroencephalographic (EEG) recordings to test the common hypothesis that this component is generated in the caudal anterior cingulate cortex. The EEG results indicated that our paradigm, a time estimation task with trial-to-trial performance feedback, elicited a large feedback-related negativity (FRN). Nevertheless, the fMRI results did not reveal any area in the caudal anterior cingulate cortex that was differentially activated by positive and negative performance feedback, casting doubt on the notion that the FRN is generated in this brain region. In contrast, we found a number of brain areas outside the posterior medial frontal cortex that were activated more strongly by positive feedback than by negative feedback. These included areas in the rostral anterior cingulate cortex, posterior cingulate cortex, right superior frontal gyrus, and striatum. An anatomically constrained source model assuming equivalent dipole generators in the rostral anterior cingulate, posterior cingulate, and right superior frontal gyrus produced a simulated scalp distribution that corresponded closely to the observed scalp distribution of the FRN. These results support a new hypothesis regarding the neural generators of the FRN, and have important implications for the use of this component as an electrophysiological index of performance monitoring and reward processing.

Introduction

An important challenge for the cognitive system is to rapidly determine the motivational significance of ongoing events. Studies using electroencephalographic (EEG) recordings have identified an event-related brain potential (ERP) correlate of this evaluative function: the feedback-related negativity (FRN), a negative deflection over (fronto-) central scalp locations peaking 250–300 ms after performance feedback (Miltner *et al*., 1997). The FRN has been studied in simple learning tasks and monetary gambling games, and is larger in amplitude for feedback stimuli associated with unfavourable outcomes (e.g. indicating erroneous performance or financial penalty) than for positive feedback (Nieuwenhuis *et al*., 2004a). The current study was designed to investigate a specific hypothesis regarding the neural generator of the FRN, and to explore possible alternative hypotheses.

Theories that have attempted to associate the FRN with specific evaluative functions have generally assumed that the FRN is generated in the caudal anterior cingulate cortex (cACC; Gehring & Willoughby, 2002; Holroyd & Coles, 2002), a brain area involved in performance monitoring (Ridderinkhof *et al*., 2004). However, although neurophysiological considerations are generally consistent with this assumption (Holroyd & Coles, 2002), direct empirical evidence for a cACC generator of the FRN is limited. First, although dipole source modelling studies have generally indicated the cACC as the most likely source of the FRN (Miltner *et al*., 1997; Gehring & Willoughby, 2002), such evidence must be interpreted with caution because of the EEG inverse problem. Second, although some functional magnetic resonance imaging (fMRI) studies have shown increased cACC activity to negative performance feedback (Ullsperger & Von Cramon, 2003; Holroyd *et al*., 2004c), other studies have failed to replicate this result (e.g. Cools *et al*., 2002).

Perhaps the most impressive evidence against a cACC source of the FRN was reported by Van Veen *et al*. (2004). They measured fMRI responses to performance feedback in a time estimation task. On each trial participants had to estimate the duration of 1 s, and were then given feedback about the quality of their estimation. Importantly, ERP studies using this task have reported large FRNs to negative feedback (Miltner *et al*., 1997; Mars *et al*., 2004). Van Veen and colleagues found no evidence for increased cACC activity to negative feedback. Interestingly, although a number of brain areas showed greater activity to positive feedback, not a single area was more activated by negative feedback than by positive feedback.

Although potentially important, the study of Van Veen *et al*. (2004) suffers from a number of limitations. First, to allow efficient deconvolution of the haemodynamic signal, the participant's response and corresponding feedback were separated by more than 10 s, compared with 1 s in ERP studies. It is possible that delaying feedback decreases its motivational significance, hence resulting in reduced cACC activity. Second, it is unclear whether the employed modified version of the task would elicit a FRN. and third, the authors did not address the question where the FRN might be generated, if not

Correspondence: Dr Sander Nieuwenhuis, as above.
E-mail: stn20@dds.nl

Received 6 January 2005, revised 8 April 2005, accepted 14 April 2005

doi:10.1111/j.1460-9568.2005.04152.x

in the cACC. In the present study, we replicated the Van Veen *et al.* experiment, while addressing these limitations.

Materials and methods

Participants

Participants in the fMRI experiment were 14 young adults (13 females, 1 male), ranging in age from 19 to 28 years (average, 21.9). All but one of these participants were right-handed and all had normal or corrected-to-normal visual acuity. Participants in a control EEG experiment were eight young adults (six females, two males), ranging in age from 19 to 25 years (average, 2.8), none of whom had taken part in the fMRI experiment. All but one of these participants were right-handed and all had normal or corrected-to-normal visual acuity. All participants were paid 15 Euros for a 1.5-h session. Written informed consent was obtained from all participants, and the experiment was approved by the research ethics committee of the Vrije Universiteit Medical Center.

Task

Identical tasks and stimuli were used for the fMRI and EEG experiments, except when noted otherwise.

Each trial started with the presentation of a visual cue, presented in the centre of the screen for 250 ms, and followed by a blank screen. The participants' task was to estimate the duration of 1 s by pressing a button as soon as they thought 1 s had elapsed following the onset of the cue. Two s following cue onset, participants received visual feedback about the accuracy of their time estimations. This resulted in ~1 s intervals between response and feedback, similar as in previous ERP studies of the time estimation task (Miltner *et al.*, 1997; Mars *et al.*, 2004). Feedback stimuli were '+' for correct estimations, '–' for incorrect estimations, and '?' in case of uninformative feedback. On each trial, it was determined randomly whether the participant received informative feedback (50%) or uninformative feedback (50%). Uninformative feedback was classified as $\text{uninformative}_{\text{correct}}$ for correct estimations or $\text{uninformative}_{\text{incorrect}}$ for incorrect estimations, and was included to control the fMRI contrast of interest (i.e. negative – positive) for haemodynamic activity associated with stimulus events and cognitive processing before feedback presentation (see Results section). The feedback stayed on the screen for 1 s, and was then followed by an intertrial interval that varied between three values occurring with roughly equal frequency: 3; 5.5; and 8 s. The interval between the feedback and the next visual cue was jittered in order to allow more efficient deconvolution of the haemodynamic signal (Burock *et al.*, 1998). Participants received instructions and 20 practice trials outside the scanner before entering the experimental phase. The experimental phase consisted of 168 trials altogether, divided into four equal blocks, with short breaks in between.

The accuracy of the time estimations was a function of whether the participants' estimates fell within a time window centred around 1 s. Unbeknownst to the participants, the width of the time window was adjusted from trial to trial, using a staircase tracking procedure (see Miltner *et al.*, 1997; for details), so that the global probability of positive and negative feedback stimuli was 50%. Participants were encouraged to try to obtain as much positive feedback as possible. After they completed the task, participants were asked to give subjective ratings of their interest in the task (1 = 'very boring'; 7 = 'very interesting'), and of how much they wanted to receive positive feedback (1 = 'indifferent'; 7 = 'very much'). For the participants in the fMRI experiment, average scores on these ratings were M = 3.5 ± 1.2 (SD) and 6.0 ± 0.8, respectively, suggesting that participants were highly motivated to perform well. For the participants in the EEG experiment, average scores were M = 4.4 ± 1.2 and 5.8 ± 0.7.

Stimuli

For the fMRI experiment, stimuli were presented in colour against a black visual display projected into the scanner. The cue consisted of a purple 'X' and subtended approximately 1.3°. The feedback stimuli were presented in a yellow, 48-size, bold Arial font and subtended approximately 1.8°. For the EEG experiment, stimuli were presented in colour against a black visual display on a monitor placed at eye level at a distance 80 cm from the participant. Stimulus colours and visual angles were roughly equal in the fMRI and EEG experiments.

fMRI image acquisition

Images were collected with a 1.5-T Siemens Sonata scanner equipped with a volume head coil. Anatomical images were collected using a T1-weighted MP-RAGE sequence (TR, 2700 ms; TE, 3.95 ms; TI, 950 ms; FA, 8°; 256*160 coronal matrix; 1.0*1.0 mm in-plane resolution; 224 1.1-mm slices). Functional images were reconstructed from 20 oblique slices acquired using a T2*-weighted EPI sequence (TR, 2000 ms; TE, 60 ms; FA, 90°; 64*64 matrix; 3.0*3.0 mm in-plane resolution; 5.0-mm slices; 20% gap). Image acquisition varied systematically across trials with respect to stimulus onset, yielding an effectively higher temporal sampling rate (Miezin *et al.*, 2000). Four functional runs (186 scans each) were collected. The first two scans of each run, recorded before the longitudinal magnetization reached a steady state recovery value, were discarded.

fMRI image analysis

Data were preprocessed and analysed with BrainVoyager software (Maastricht, the Netherlands). Image preprocessing consisted of: rigid-body three-dimensional (3D) motion correction using trilinear interpolation; slice scan time correction using sinc interpolation; spatial smoothing with a 4-mm fullwidth at half maximum (FWHM) Gaussian kernel; voxel-wise linear detrending, highpass filtering (above 7 cycles per time course) to remove low frequencies, and lowpass filtering with a 2.8-s FWHM Gausian kernel to remove high frequencies. Spatial normalization was performed using the standard 9-parameter landmark method of Talairach & Tournoux (1988). For each participant, the blood oxygen-level dependent (BOLD) responses across the scanning run were modelled with a general linear model that included five regressors. Four regressors accounted for the four possible feedback types (positive, negative, $\text{uninformative}_{\text{correct}}$, $\text{uninformative}_{\text{incorrect}}$). An additional regressor accounted for the visual cue. The haemodynamic response to each event was estimated by convolving each regressor with a standard gamma function (Boynton *et al.*, 1996). For each voxel and each event type, a parameter estimate was generated that indicated the strength of covariance between the data and the haemodynamic response function; these estimates were corrected for temporal autocorrelation using a first-order autoregressive model. Contrasts between parameter estimates for different events were calculated for each participant, and the results submitted to a group analysis that treated intersubject variability as a random effect. Statistical parametric maps were derived from the resulting *t*-values associated with each voxel and were thresholded at a conservative value ($P < 0.0005$, uncorrected), with a contiguity threshold of 120 mm^3 as a further precaution against type-1

errors (Forman *et al*., 1995). The location of the peak activity associated with each cluster of activation was reported in Talairach coordinates (Talairach & Tournoux, 1988).

EEG data acquisition and analysis

EEG recordings were taken from 32 Ag/AgCl electrodes embedded in a fabric cap (Electro-Cap International, Inc., Eaton, OH, USA), and placed in an extended 10–20 system montage, referenced to the left mastoid. During offline analysis, all signals were re-referenced to the algebraic mean of both mastoids. The electro-oculogram (EOG) was recorded from electrodes placed above and below the left eye, and from electrodes placed on the outer canthi of each eye. All electrode impedances were kept below 10 kΩ. The EEG signals were amplified (Synamps, bandpass filter 0.1–70 Hz), and digitized at 250 Hz.

Single-trial epochs were extracted offline for a period from 200 ms before until 800 ms after the critical event. Standard Neuroscan (Neurosoft Inc., Sterling VA, USA) analysis procedures were used to correct for EOG artifacts and to discard trials with recording artifacts. Then, for each participant and each condition, the EEG epochs were averaged with respect to feedback onset to obtain feedback-locked ERPs. A baseline, computed as the average signal activity across the 200 ms before the feedback stimulus, was subtracted for each ERP. The resulting ERP waveforms were low-pass filtered (< 16 Hz, 12 dB/oct, zero-phase shift). Following previous studies using this paradigm (Miltner *et al*., 1997; Mars *et al*., 2004), difference waveforms were created by subtracting the signal elicited on trials with positive feedback from the signal elicited on trials with negative feedback. For each participant the amplitude of the FRN was defined as the peak negativity of the difference waveform at electrode Cz (where the FRN reached its maximum amplitude) in a window 200–350 ms following feedback onset.

BESA 2000 (MEGIS software GmbH, Gräfelfing, Germany) was used to model the scalp distribution of the FRN in the difference waveform (negative feedback − positive feedback) with a combination of equivalent dipole sources. Modelling was performed on data re-referenced to the average reference across a 16-ms window around the FRN peak (i.e. 292–308 ms), using the standard BESA four-shell spherical head model (radius, 85 mm; thickness scalp, 6 mm; thickness bone, 7 mm; and thickness cerebrospinal fluid, 1 mm). The locations of dipoles were fixed and were based on the Talairach coordinates of the areas identified by the fMRI analysis (see below). The orientations of the dipoles were unconstrained. The BESA algorithm minimized the residual variance between the scalp distribution simulated by the equivalent dipole source model and the observed FRN scalp distribution.

Results

As a consequence of the task design, each participant received negative feedback on approximately 25% of the trials, positive feedback on approximately 25% of the trials, and uninformative feedback on approximately 50% of the trials. Trials with uninformative feedback were classified as $\text{uninformative}_{\text{incorrect}}$ (∼25%) or $\text{uninformative}_{\text{correct}}$ (∼25%) according to the quality of the time estimation.

fMRI experiment

To investigate whether participants used the informative feedback stimuli to improve their performance, we scored the absolute value of the time estimation error on each trial with regard to whether the preceding trial involved informative or uninformative feedback. Participants' time estimations were more accurate following informative feedback (timing error, 54 ms) than following uninformative feedback (timing error, 69 ms; $t_{13} = 2.3$, $P < 0.05$, two-tailed), indicating that they used the feedback to improve their performance.

To identify brain areas that were activated more by negative feedback than by positive feedback, we performed the following contrast: (negative − positive) minus ($\text{uninformative}_{\text{incorrect}}$ − $\text{uninformative}_{\text{correct}}$). An analogous contrast was performed to identify brain areas that were activated more by positive feedback than by negative feedback. This double subtraction ensured that any obtained brain activation could be attributed to differential processing of negative and positive feedback, and not to potential differences in information processing on correct and incorrect trials (e.g. with regard to strength of attention or uncertainty about performance), or to differences in trial history (e.g. correct trials are generally preceded by correct trials and vice versa).

At the statistical threshold of $P < 0.0005$ (uncorrected) there were no brain areas that exhibited greater activity for negative than for positive feedback. Even lowering the threshold to $P < 0.05$ (uncorrected) revealed no areas in the cACC or adjacent regions of the posterior medial frontal cortex that were activated more by negative feedback. In contrast, various brain areas showed greater activity to positive than to negative feedback. These included areas in the rostral anterior cingulate cortex, right superior frontal gyrus, posterior cingulate cortex, striatum (caudate/putamen), and cerebellum (see Table 1 and Fig. 1).

To explore the possibility that the cACC was activated to a similar degree by positive and negative feedback, we performed a whole-brain conjuction analysis to identify areas in which the estimated BOLD response showed the following pattern: [positive feedback > baseline (i.e. fixation)] AND [negative feedback > baseline]. This analysis revealed a large area in the posterior medial frontal cortex, extending from the cACC into the presupplementary motor area ($x = -1$, $y = 0$, $z = 54$; Fig. 2A). Examination of the event-related BOLD signal averages associated with this area indicated essentially overlapping BOLD signal increases for the various types of feedback (Fig. 2B). Interestingly, these signal increases started slightly before feedback onset, suggesting that the posterior medial frontal cortex was activated not by the feedback but by an internal response evaluation process, anticipation of the feedback, or perhaps the visual cue.

TABLE 1. Brain areas showing greater activity for positive feedback than for negative feedback

		Talairach coordinates			
Area	Left/Right	*x*	*y*	*z*	Max *t*-value
Rostral anterior cingulate cortex	Right	3	43	16	5.56
Rostral anterior cingulate cortex	–	0	40	−2	9.13
Superior frontal gyrus	Right	20	39	43	6.51
Posterior cingulate cortex	Left	−1	−30	33	7.42
Caudate/putamen	Left	−14	8	5	7.57
Caudate/putamen	Right	16	7	4	8.55
Cerebellum	Left	−31	−70	−22	5.87

Note that all regions are $P < 0.0005$ (uncorrected, voxel contiguity = 120 mm^3).

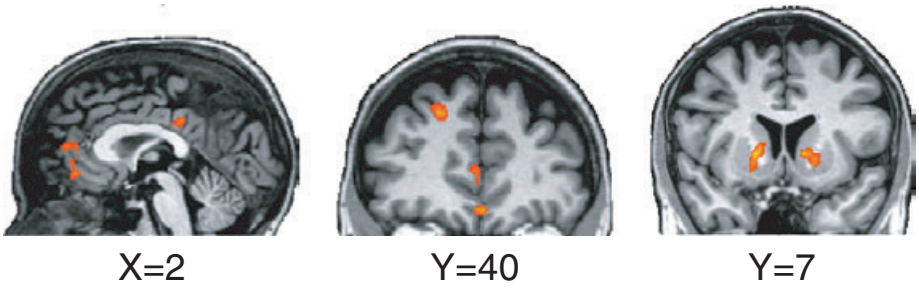

FIG. 1. Brain areas showing greater activity for positive feedback than for negative feedback. $P < 0.0005$ (uncorrected, voxel contiguity = 120 mm^3). Left: activations in the rostral anterior cingulate cortex (two foci) and posterior cingulate cortex. Middle: activations in the rostral anterior cingulate cortex (two foci) and right superior frontal gyrus. Right: activations in the left and right caudate/putamen.

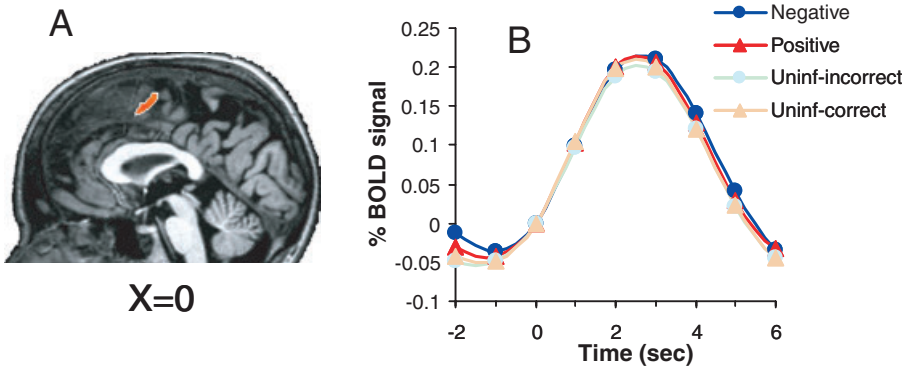

FIG. 2. (A) Region of the cACC/presupplementary motor area that is activated to a similar extent by positive and feedback, as revealed by a conjunction analysis (see text for details). (B) Event-related BOLD signal averages associated with this area. Note that the BOLD signal responses set off before the presentation of the feedback, suggesting that they are elicited by an event preceding the feedback.

EEG experiment

The time estimation task used in the fMRI experiment differed in two important aspects from the task design used in ERP studies (Miltner *et al.*, 1997; Mars *et al.*, 2004). These differences involved the use of long and variable intertrial intervals and the use of uninformative feedback. One aim of the present EEG experiment was to verify whether our modified task would yield similar FRNs as in previous ERP studies.

Figure 3 (upper panel) shows the ERPs elicited by negative, positive, and uninformative feedback stimuli. As can be seen, negative feedback was associated with a negative deflection (superimposed on the P3) between ~200–350 ms after the feedback stimulus, replicating previous ERP studies. The scalp map in Fig. 4A shows that the FRN had a central scalp distribution that was slightly lateralized to the right of the midline. The ERPs associated with uninformative feedback were markedly different from the ERPs associated with negative and positive feedback, showing a pronounced N2, perhaps reflecting the uncertainty associated with the uninformative feedback, and a smaller P3 (cf. Müller *et al.*, 2005). Figure 3 (lower panel) shows, for each individual, the difference wave obtained by subtracting the ERP associated with positive feedback from the ERP associated with negative feedback. Figure 3 demonstrates that the FRN was consistently observed in the current task; seven out of eight participants exhibited a clear FRN. The average peak latency of the FRN was M = 300 ± 26 ms (SD). The average peak amplitude across individuals was M = −9.4 ± 4.4 μV.

Dipole source modelling

We used a neuroanatomically constrained dipole source analysis to investigate how well activation of the areas identified by the fMRI analysis (Table 1) could account for the observed scalp distribution of the FRN (see Materials and methods for details). We focused on the four cortical areas (two in rostral ACC, right superior frontal gyrus, and posterior cingulate cortex), as activation of subcortical areas is unlikely to contribute significantly to the scalp-recorded EEG. We computed and compared the goodness-of-fit of each of various combinations of dipoles seeded in these four cortical areas. None of the single-dipole models yielded a satisfactory fit [all residual variances (RVs) > 10%]. Of all possible two-dipole models, the combination of the most ventral rostral ACC source and the posterior cingulate source resulted in the best fit (RV = 5.6%; see Fig. 4B). Of all possible three-dipole models, the above combination together with a source in the right superior frontal gyrus explained most of the variance (RV = 4.4%; Fig. 4C). A model with dipoles in all four cortical areas yielded essentially no improvement in fit (RV = 4.3%). Finally, we examined one- and two-dipole models in which both the location and orientation of the dipoles were unconstrained. Irrespective of the dipole starting locations, this led to a solution with one dipole in a deep and implausible location outside of the brain.

Discussion

The aim of this study was to investigate the neural correlates of processing performance feedback. More specifically, we were inter-

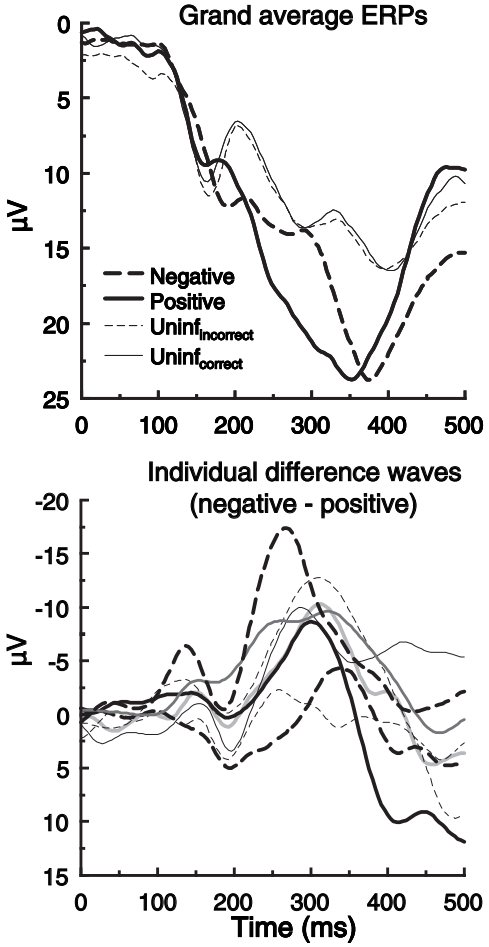

FIG. 3. Results from the EEG experiment. Upper panel: Grand-average ERP waveforms from electrode Cz for each of the four feedback conditions. Uninf, uninformative. Lower panel: ERP difference waves (negative–positive feedback) for each of the eight participants.

ested in determining the neural generators of the FRN, an ERP component that is modulated by the valence of performance feedback (Miltner *et al*., 1997; Yeung & Sanfey, 2004). The existing evidence, though mostly indirect, has led to the proposal that the FRN is generated in the cACC or a closely adjacent area in the posterior medial frontal cortex (Miltner *et al*., 1997; Gehring & Willoughby, 2002; Holroyd & Coles, 2002; Nieuwenhuis *et al*., 2004a), an important area for performance monitoring and reward processing (Ridderinkhof *et al*., 2004; Rushworth *et al*., 2004). In the current study we tested this proposal by measuring fMRI signals from participants performing a time estimation task with trial-to-trial performance feedback, an experimental paradigm often used to study the FRN (e.g. Miltner *et al*., 1997). Importantly, we found no areas in or near the cACC in which activity was modulated by the valence of the feedback. In contrast, we found a number of brain areas outside the posterior medial frontal cortex that were activated more strongly by positive feedback than by negative feedback. These included areas in the rostral ACC, right superior frontal gyrus, posterior cingulate cortex, and striatum. These results cast doubt on the notion that the FRN is generated in the cACC and instead suggest other neural generators.

In a recent fMRI study using the same task, Van Veen and colleagues found essentially the same results as reported here (Van Veen *et al*., 2004). Our reasons for replicating this study were threefold. First, in the study of Van Veen *et al*. the participant's response and corresponding feedback were separated by more than 10 s, compared to 1 s in ERP studies (e.g. Miltner *et al*., 1997). We hypothesized that this long delay may have decreased the motivational significance of the feedback, which therefore may have failed to activate the cACC. In the current study the response−feedback interval was 1 s; the use of an uninformative feedback condition allowed us to dissociate brain activity associated with feedback valence from response related brain-activity (i.e. correct vs. incorrect time estimations) and brain activity on the preceding trial. Second, unlike Van Veen and colleagues, we performed a control EEG experiment to demonstrate that the FRN is consistently observed in our version of the time estimation task, even though some of the task parameters were different than in previous ERP studies, and third, to address the question where the FRN might be generated, if not in the cACC, we modelled the observed scalp distribution of the FRN using dipole source models that were anatomically constrained by the results from the fMRI experiment. For these reasons, our study is probably the most thorough investigation thus far of the source of the feedback-related negativity.

The dipole source analyses indicated that the FRN scalp distribution could be explained reasonably well by a two-dipole model with dipoles in the posterior cingulate and ventral rostral ACC, two areas that showed differential fMRI activity to positive and negative feedback (see also Van Veen *et al*., 2004). Interestingly, this model is very similar to an FRN dipole model that has recently been proposed by Müller *et al*. (2005; see also Luu *et al*., 2003). These authors examined a two-dipole model that was anatomically constrained by the results of a meta-analysis of fMRI studies of self evaluation, a function that is presumably engaged by the delivery of performance feedback. Forward modelling indicated that activity in the two primary brain areas identified by this meta-analysis, the posterior cingulate and medial prefrontal cortex (slightly more anterior and dorsal than our rostral ACC activations), could explain most of the variance of the FRN scalp distribution observed by Müller and colleagues, thus providing converging evidence for our two-dipole model. A characteristic property of the FRN scalp distribution in the current study and in some previous studies (Gehring & Willoughby, 2004; Nieuwenhuis *et al*., 2004b; Yeung & Sanfey, 2004) was that it was lateralized slightly to the right of the midline. The fMRI results suggest that this may be explained by activity in the right superior frontal gyrus, which was more pronounced following positive feedback than following negative feedback. This hypothesis received support from our dipole source analyses: Extending the above discussed two-dipole model with a third dipole in the right superior frontal gyrus led to a close correspondence between the observed and simulated scalp distributions (see Fig. 4), and a further increase in explained variance. Taken together, the fMRI results and the dipole source analyses suggest that the FRN reflects the summed activity of generators in the posterior cingulate, the rostral ACC, and (in some experiments) the right superior frontal gyrus.

The sensitivity of the posterior cingulate and rostral ACC to feedback valence seems consistent with existing literature on the anatomical properties and functional significance of these brain areas. The posterior cingulate is interconnected with reward-related areas of the brain (including the ACC, orbitofrontal cortex, and caudate nucleus), and is activated by the expectation and delivery of reward, at least following oculomotor responses (McCoy *et al*., 2003). Neural activity in the posterior cingulate also scales with the difference between expected and actual reward (i.e. a reward prediction error), consistent with similar findings for the FRN (Holroyd *et al*., 2004a).

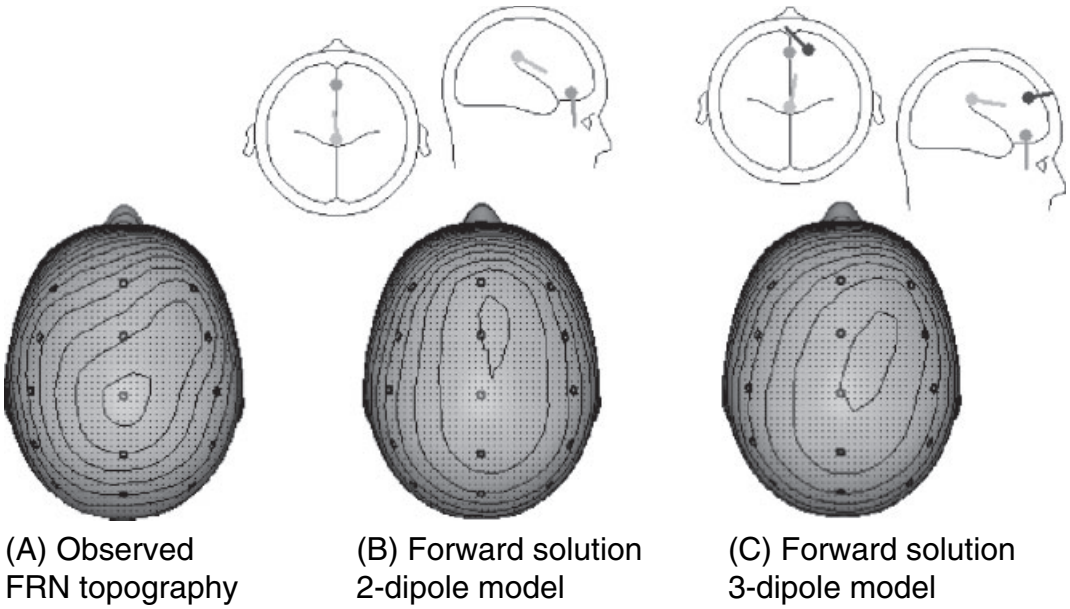

FIG. 4. (A) Grand-average, average-reference, spline-interpolated iso-potential map reflecting the scalp topography of the FRN in the negative–positive difference wave at its peak latency, t = 300 ms. (B) Simulated scalp topography (forward solution) associated with an anatomically constrained model with dipoles in the ventral rostral ACC and posterior cingulate. (C) Simulated scalp topography associated with an anatomically constrained model with dipoles in the ventral rostral ACC, posterior cingulate, and right superior frontal gyrus. The difference in voltage value represented by each isopotential line is 0.4 μV.

The rostral ACC appears primarily involved in assessing the salience of emotional and motivational information, and in regulating emotional responses (Bush *et al*., 2000). For example, the rostral ACC is one of the brain areas that is typically activated following error responses in speeded response tasks (Kiehl *et al*., 2000).

Although we did not find any evidence that the FRN is generated in the cACC, this possibility cannot be entirely excluded on the basis of our fMRI results. fMRI BOLD and EEG signals originate from distinct physiological mechanisms. EEG reflects the summation of excitatory and inhibitory postsynaptic potentials in pyramidal cells, and is largely independent of action potentials. In contrast, fMRI reflects changes in metabolic demands related to activity of all neural cells, and is dependent on local field potentials (Logothetis, 2003). Therefore, one can easily imagine a region of cortical tissue with large metabolic load making little contribution to the EEG. The opposite pattern, strong EEG signals but weak metabolic signatures, is also imaginable. For example, synchronous activity of only a few percent of the neurons in each cortical column, but in the context of large-scale synchrony among different columns, would produce a large electric scalp field, but minimal metabolic signatures (Nunez & Silberstein, 2000). In this way, a phasic performance-related error signal carried by the midbrain dopamine system (Schultz, 2002; Aron *et al*., 2004) might generate the FRN by modulating cortical activity across a wide neural area in the frontal midline (Holroyd & Coles, 2002).

In addition to the possibility of a dissociation between fMRI and EEG, we considered the possibility that the cACC was activated to a similar extent by positive and negative feedback, and therefore was not identified by our main contrast. Note that each of the areas identified by the fMRI analysis showed a greater activation to positive feedback than to negative feedback, whereas the FRN is largest following negative feedback. This apparent contradiction can be resolved by assuming that: (i) there is a brain area that is equally activated by positive and negative feedback, and the activity of which causes a 'baseline' negativity over central scalp locations (possibly superimposed on the P3); and (ii), there are one or more brain areas that are differentially activated by positive feedback, and whose activity modulates the baseline negativity in positive direction, yielding a less negative deflection following positive feedback (cf. Van Veen *et al*., 2004). The general possibility that the FRN reflects a modulation of a baseline negative ERP component (the N2) has been discussed in detail by Holroyd (2004). In the current research we examined whether the cACC might be a candidate for generating this baseline negativity, in which case it should show similar activation following positive and negative feedback. Our exploratory analysis suggested that this was not the case: Although we identified a region of the cACC (extending into the presupplementary motor area) that was equally activated on all trials, irrespective of the type of feedback, the BOLD signal time courses revealed that this area was activated already before the feedback, possibly by a process of internal response evaluation. This would be consistent with previous findings of cACC activation in the context of uncertain or underdetermined responding (cf. Holroyd *et al*., 2004c; Ridderinkhof *et al*., 2004). Note that we cannot rule out the possibility that the cACC is also activated to some extent by the feedback stimuli, and that this activity sums with the activity caused by events preceding the feedback. Alternatively, a baseline negativity might be generated in areas outside the cACC.

Although our results are consistent with some previous fMRI studies that have failed to find sensitivity of the cACC to the valence of abstract performance feedback (Cools *et al*., 2002; Aron *et al*., 2004; Van Veen *et al*., 2004), other studies have reported increased cACC activity to negative feedback (Monchi *et al*., 2001; Ullsperger & Von Cramon, 2003; Holroyd *et al*., 2004c). Similarly mixed results have been obtained by studies that have used monetary rewards and punishments instead of abstract performance feedback (e.g. Bush *et al*., 2002; Nieuwenhuis *et al*., 2005). The discrepancy between these results can potentially be explained in terms of the relative frequency of unfavourable (e.g. negative feedback and monetary punishments) and favourable outcomes. In studies that have reported reliable cACC

activation in association with unfavourable outcomes, these outcomes generally occurred infrequently compared to favourable outcomes. In these studies the cACC activation seems to reflect an interaction between valence and frequency rather than an effect of frequency per se, as infrequent control stimuli without valence did not reliably activate the cACC (Bush *et al*., 2002; Ullsperger & Von Cramon, 2003). Conversely, in the studies that have failed to find reliable cACC activation in association with unfavourable outcomes, unfavourable and favourable outcomes generally occurred with equal frequency (but see Cools *et al*., 2002). Importantly, FRN amplitude is inversely related to the frequency of unfavourable outcomes (Nieuwenhuis *et al*., 2002; Holroyd *et al*., 2003), suggesting that task designs with a lower frequency of unfavourable outcomes may have higher power to detect the brain activity underlying the FRN. Nonetheless, the large average FRN amplitude found in the present study ($\sim$10 μV) is inconsistent with the possibility that our task design may have had insufficient power to detect the source of the FRN.

If the present findings are correct in suggesting that the FRN is not generated in the cACC, this would have implications for existing theories of the functional significance of the FRN, and in particular for the reinforcement learning theory of Holroyd & Coles (2002). This theory attributes a critical role to the dorsal cACC in reward-based learning, using various kinds of evaluative information for the reinforcement of adaptive behaviours. Furthermore, the theory holds that unexpected negative and positive events differentially modulate cACC activity. These tenets of the theory have received strong support from human neuroimaging studies and neurophysiological recording studies (e.g. Bush *et al*., 2002; Ito *et al*., 2003; Williams *et al*., 2004). In addition, the theory claims that the cACC responses to errors and unfavourable outcomes are reflected in two similar electrophysiological scalp potentials, the FRN and the error-related negativity (ERN), a negative potential peaking approximately 80 ms following erroneous responses (for review see Holroyd *et al*., 2004b). However, although the FRN and ERN have many properties in common (Holroyd & Coles, 2002), the current results challenge an important prediction of the reinforcement learning theory, namely that the two ERP components are generated in the same brain area: Whereas erroneous responses and the ERN are robustly associated with activation of the cACC (Holroyd *et al*., 2004b), the evidence for a cACC generator of the FRN is mixed.

To conclude, we have presented evidence against the common hypothesis that the FRN indexes the response of the cACC to unfavourable outcomes. Instead, our fMRI results along with the results from EEG source analyses suggest that the FRN reflects the summed activity of regions in the rostral ACC and the posterior cingulate (and in some experiments the right superior frontal gyrus), replicating and extending recent studies by Van Veen *et al*. (2004) and Müller *et al*. (2005). These results provide important information for the study of human performance monitoring and reward processing, and in particular for the interpretation of the growing literature on the FRN.

Acknowledgements

This research was supported by the Netherlands Organization for Scientific Research (S.N.).

Abbreviations

BOLD, blood oxygen-level dependent; cACC, caudal anterior cingulate cortex; EEG, electroencephalography; ERP, event-related brain potential; fMRI, functional magnetic resonance imaging; FRN, feedback-related negativity.

References

Aron, A.R., Shohamy, D., Clark, J., Myers, C., Gluck, M.A. & Poldrack, R.A. (2004) Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *J. Neurophysiol.*, **92**, 1144–1152.

Boynton, G.M., Engel, S.A., Glover, G.H. & Heeger, D.J. (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.*, **16**, 4207–4221.

Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R. & Dale, A.M. (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport*, **9**, 3735–3739.

Bush, G., Luu, P. & Posner, M.I. (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.*, **4**, 215–222.

Bush, G., Vogt, B.A., Holmes, J., Dale, A.M., Greve, D., Jenike, M.A. & Rosen, B.R. (2002) Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proc. Natl. Acad. Sci. USA*, **99**, 523–528.

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. *J. Neurosci.*, **22**, 4563–4567.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A. & Noll, D.C. (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.*, **33**, 636–647.

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

Gehring, W.J. & Willoughby, A.R. (2004) Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. In Ullsperger, M. & Falkenstein, M. (Eds), *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring.* MPI of Cognitive Neuroscience, Leipzig, pp. 14–20.

Holroyd, C.B. (2004) *A note on the oddball N200 and the feedback ERN*. In Ullsperger, M. & Falkenstein, M. (Eds), *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring.* MPI of Cognitive Neuroscience, Leipzig, pp. 211–218.

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

Holroyd, C.B., Larsen, J.T. & Cohen, J.D. (2004a) Context dependence of the event-related brain potential to reward and punishment. *Psychophysiology.*, **41**, 245–253.

Holroyd, C.B., Nieuwenhuis, S., Mars, R.B. & Coles, M.G.H. (2004b) Anterior cingulate cortex, selection for action, and error processing. In Posner, M.I. (Ed.), *The Cognitive Neuroscience of Attention*. Guilford Press, New York, pp. 219–231.

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.

Holroyd, C.B., Nieuwenhuis, S., Yeung, N., Nystrom, L.E., Mars, R.B., Coles, M.G.H. & Cohen, J.D. (2004c) Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nat. Neurosci.*, **7**, 497–498.

Ito, S., Stuphorn, V., Brown, J.W. & Schall, J.D. (2003) Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*, **302**, 120–122.

Kiehl, K.A., Liddle, P.F. & Hopfinger, J.B. (2000) Error processing and the rostral anterior cingulate: an event-related fMRI study. *Psychophysiology*, **37**, 216–223.

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

Logothetis, N.K. (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. *J. Neurosci.*, **23**, 3963–3971.

Mars, R.B., De Bruijn, E.R.A., Hulstijn, W., Miltner, W.H.R. & Coles, M.G.H. (2004) What if I told you 'you were wrong'? Brain potentials and behavioral adjustments elicited by performance feedback in a time-estimation task. In Ullsperger, M. & Falkenstein, M. (Eds), *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring.* MPI of Cognitive Neuroscience, Leipzig, pp. 129–134.

McCoy, A.N., Crowley, J.C., Haghighian, G., Dean, H.L. & Platt, M.L. (2003) Saccade reward signals in posterior cingulate cortex. *Neuron*, **40**, 1031–1040.

Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E. & Buckner, R.L. (2000) Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*, **11**, 735–759.

Miltner, W.H.R., Braun, C.H. & Coles, M.G.H. (1997) Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a generic neural system for error detection. *J. Cogn. Neurosci.*, **9**, 788–798.

Monchi, O., Petrides, M., Petre, V., Worsley, K. & Dagher, A. (2001) Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J. Neurosci.*, **21**, 7733–7741.

Müller, S.V., Möller, J., Rodriguez-Fornells, A. & Münte, T.F. (2005) Brain potentials related to self-generated and external information used for performance monitoring. *Clin. Neurophys.*, **116**, 63–74.

Nieuwenhuis, S., Heslenfeld, D.J., Alting von Geusau, N.J., Mars, R.B., Holroyd, C.B. & Yeung, N. (2005) Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage*, **25**, 1302–1309.

Nieuwenhuis, S., Holroyd, C.B., Mol, N. & Coles, M.G.H. (2004a) Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neurosci. Biobehav. Rev.*, **28**, 441–448.

Nieuwenhuis, S., Ridderinkhof, K.R., Talsma, D., Coles, M.G.H., Holroyd, C.B., Kok, A. & Van der Molen, M.W. (2002) A computational account of altered error processing in older age: Dopamine and the error-related negativity. *Cogn. Affect. Behav. Neurosci.*, **2**, 19–36.

Nieuwenhuis, S., Yeung, N., Holroyd, C.B., Schurger, A. & Cohen, J.D. (2004b) Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cereb. Cortex*, **14**, 741–747.

Nunez, P.L. & Silberstein, R.B. (2000) On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense? *Brain Topogr.*, **13**, 79–96.

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.

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

Schultz, W. (2002) Getting formal with dopamine and reward. *Neuron*, **36**, 241–263.

Talairach, J. & Tournoux, P. (1988) *Co-Planar Stereotaxic Atlas of the Human Brain: an Approach to Medical Cerebral Imaging*. Thieme, Stuttgart, Germany.

Ullsperger, M. & Von Cramon, D.Y. (2003) Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *J. Neurosci.*, **23**, 4308–4314.

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. Cogn.*, **56**, 267–276.

Williams, Z.M., Bush, G., Rauch, S.L., Cosgrove, C.R. & Eskandar, E.N. (2004) Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat. Neurosci.*, **7**, 1370–1375.

Yeung, N. & Sanfey, A.G. (2004) Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.*, **24**, 6258–6264.