

Separable neural mechanisms contribute to feedback processing in a rule-learning task

K. Zanolie^{a,b,c}, L. Van Leijenhorst^{a,b,c}, S.A.R.B. Rombouts^{a,b,c}, E.A. Crone^{a,b,c,*}

^a *Leiden University Institute for Psychological Research (LU-IPR), Leiden University, The Netherlands*

^b *Leiden Institute for Brain & Cognition (LIBC), Leiden, The Netherlands*

^c *Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands*

Received 22 February 2007; received in revised form 2 August 2007; accepted 6 August 2007

Available online 22 August 2007

Abstract

To adjust performance appropriately to environmental demands, it is important to monitor ongoing action and process performance feedback for possible errors. In this study, we used fMRI to test whether medial prefrontal cortex (PFC)/anterior cingulate cortex (ACC) and dorsolateral (DL) PFC have different roles in feedback processing. Twenty adults completed a rule-switch task in which rules had to be inferred on the basis of positive and negative feedback and the rules could change unexpectedly. Negative feedback resulted in increased activation in medial PFC/ACC and DLPFC relative to positive feedback, but the regions were differentially active depending on the type of negative feedback. Whereas medial PFC/ACC was most active following unexpected feedback indicating that prior performance was no longer correct, DLPFC was most active following negative feedback that was informative for correct behavior on the next trial. The current findings show that inconsistent results about the role of prefrontal cortex regions in feedback processing are most likely associated with the informative value of the performance feedback. The results are consistent with the hypothesis that medial PFC/ACC is important for signaling expectation violation whereas DLPFC is important for goal-directed actions. © 2007 Elsevier Ltd. All rights reserved.

Keywords: Prefrontal cortex; Feedback; Performance monitoring; fMRI; WCST

1. Introduction

An important function of adaptive control is using feedback to monitor our own performance for possible errors and make rapid adjustments. This type of performance monitoring has been well documented in the classic Wisconsin card sorting task (WCST) (Demakis, 2003; Miltner, 1963). In this task, participants must sort cards according to different possible sorting rules, related to color, shape or number. Feedback after each sorting trial indicates whether performance was correct or incorrect. Once the participant has discovered the correct rule, the rule changes unexpectedly and the participant needs to use positive and negative feedback cues to find the new correct sorting rule.

An important feature of the WCST is the processing of negative feedback (Bayless et al., 2006; Monchi, Petrides, Petre, Worsley, & Dagher, 2001), which can have several meanings

(Barcelo & Knight, 2002). First, negative feedback that indicates that previous performance is no longer correct can be referred to as first-warning negative feedback. This scenario entails a violation of expectation due to the unexpected nature of the changes in the rule that has to be used (Monchi et al., 2001). Second, negative feedback that can be used to test hypotheses about the new sorting rule can be referred to as efficient negative feedback. In this case, the feedback indicates that possibilities for future actions are constrained, and that the participant can use goal-directed behavior to apply the correct rule on the next trial (Walton, Devlin, & Rushworth, 2004). Third, negative feedback can indicate that a performance error was committed. This feedback signals an internal violation of expectations and also requires the participant to exert goal-directed behavior on the next trial (Holroyd & Coles, 2002).

Neuroimaging studies may allow us to examine the possible dissociability of underlying mechanisms for these different types of negative performance feedback. To date, most neuroimaging studies have focused on the differences in neural activity associated with negative relative to positive feedback (Holroyd et al., 2004; van Veen, Holroyd, Cohen, Stenger, & Carter, 2004),

* Corresponding author at: Department of Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.

E-mail address: ecrone@fsw.leidenuniv.nl (E.A. Crone).

but the results are inconclusive (Nieuwenhuis, Slagter, Alting von Geusau, Heslenfeld, & Holroyd, 2005). These studies have not attempted to dissociate between different types of feedback processing, and a review suggests that the performance feedback types summarized above may indeed be neurally separable.

Event-related potential (ERP) and source localization studies have suggested that the medial prefrontal cortex (PFC) is important for processing errors and negative feedback (Holroyd & Coles, 2002; Miltner, 1963; Nieuwenhuis, Holroyd, Mol, & Coles, 2004). The feedback-related negativity (FRN) is a negative deflection over (fronto-) central scalp locations originating in or near medial PFC. It peaks 250–300 ms after negative performance feedback and is thought to correlate with the evaluative function of the feedback (Gehring & Willoughby, 2002; Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997; Nieuwenhuis et al., 2004). Studies using functional magnetic resonance imaging (fMRI) have confirmed that medial PFC, mostly the anterior cingulate cortex (ACC), is active following errors and error feedback (Holroyd et al., 2004; Mars et al., 2005; Ullsperger & von Cramon, 2003). The finding that the medial PFC/ACC is important for processing errors and negative feedback suggests that this region provides the first indication of outcomes worse than anticipated (Brown and Braver, 2005; Holroyd et al., 2004; Kerns et al., 2004; Rushworth, Walton, Kennerley, & Bannerman, 2004). However, other neuroimaging studies have not consistently shown that medial PFC/ACC differentiates between negative and positive feedback (Nieuwenhuis et al., 2005; van Veen et al., 2004).

Lateral prefrontal cortex (lateral PFC) has also been implicated in processing negative performance feedback. For example, set-shifting studies that have used paradigms closely matching the WCST have consistently reported activation in dorsolateral (DL) PFC following switch cues (Dove, Pollmann, Schubert, Wiggins, & Von Cramon, 2000; Konishi, Hayashi, Kikyo, Takahashi, & Miyashita, 2002; Lie, Specht, Marshall, & Fink, 2006). In addition, Dias, Robbins, and Roberts (1997) reported that damage to lateral PFC in marmosets results in impaired performance on extradimensional switch trials.

Lateral PFC, including mid-DLPFC, is generally thought to be involved in monitoring of task sets in working memory (Kerns et al., 2004; Monchi et al., 2001). A feedback-learning study by Monchi et al. (2001) indicated that ventrolateral PFC is more active following negative feedback compared to positive feedback. In contrast, Nieuwenhuis et al. (2005) found in an fMRI study and ERP dipole source analyses that the right superior frontal gyrus was activated more strongly by positive than by negative feedback. Thus, regions within lateral PFC can be sensitive to both negative and positive feedback. It is possible that regions within lateral PFC are sensitive to the informative value of the feedback, rather than the valence of the feedback. This hypothesis finds support in studies showing that lateral PFC damage in humans impairs corrective behavior. Thus, the lateral PFC possibly interacts with medial PFC/ACC in monitoring behavior (Gehring & Knight, 2000).

Besides medial PFC/ACC and DLPFC, lateral orbitofrontal cortex (lat-OFC, area BA 47) (O'Doherty, Critchley,

Deichmann, & Dolan, 2003) and the caudate nucleus (Monchi et al., 2001; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006) have also been implicated in the processing of feedback in set-shifting tasks such as the WCST. Whereas lat-OFC has been implicated in more emotionally salient feedback processing (Dias et al., 1997; O'Doherty et al., 2003), the caudate has been found to be active following both positive and negative feedback (Tricomi et al., 2006). The latter result has been interpreted to mean that the caudate is involved in the general process of learning, but it remains unclear whether this region is sensitive to the informative value of the feedback.

The current study examined the role of the medial PFC/ACC and mid-DLPFC using a rule-learning task based on the classic Wisconsin card sorting task (WCST). Participants were presented with a stimulus display containing four horizontally spaced locations. A stimulus was randomly placed in front of one of the four locations. Participants were instructed to apply one of three stimulus–response mapping rules that had been practiced prior to scanning. After applying a rule, participants received positive or negative performance feedback (Crone, Ridderinkhof, Worm, Somsen, & van der Molen, 2004). The location task has two advantages over the classic WCST. First, all rules were based on the same dimension (location), eliminating the demands for extradimensional switches. Second, in the classic WCST some stimuli can be sorted according to multiple rules (e.g. color and shape), which can lead to ambiguous trials. This possibility was eliminated in the present design, because each location rule was specifically mapped to one response (see also Walton et al., 2004). During scanning, participants were not told which rule to apply and were instructed to use the positive or negative feedback in order to deduce the correct rule. After two, three or four correct consecutive rule applications the rule changed without warning, and participants were instructed to use positive and negative feedback to find the correct rule again.

We differentiated between three types of negative feedback, each of different informative value for the participant when performing the task. The first type of negative feedback (negFB) is the first-warning negFB, which indicates a rule-switch. An efficient negFB indicates that the rule chosen when searching for the correct rule is not correct. Therefore, the participant knows with certainty what the correct rule should be on the next trial. An error is only scored as efficient negative feedback when it is preceded by a first-warning negFB and followed by a correctly applied rule. The third type of negative feedback is the error-related negFB, which indicates that the participant failed to use the correct rule in a set of correct trials.

Based on prior studies indicating that medial PFC/ACC is active when there is high conflict (Mars et al., 2005) or a large number of response alternatives (Walton et al., 2004), we predicted increased activation in this region relative to DLPFC following a first-warning negFB. In contrast, we expected that DLPFC would be more active than medial PFC/ACC following an efficient negFB, because there are fewer response alternatives and the demand for goal-directed behavior is higher (Miller & Cohen, 2001). Both medial PFC/ACC and DLPFC were expected to be active following error-related negFB. In addition to medial PFC/ACC and DLPFC, we also tested the role of

lat-OFC (Monchi et al., 2001; O’Doherty et al., 2003) and the caudate nucleus (Monchi et al., 2001; Tricomi et al., 2006) in feedback processing.

2. Methods

2.1. Participants

Twenty adults between 18 and 25 years of age ($M=20.6$, $S.D.=1.92$, 8 male) were recruited from Leiden University. The participants, all of whom received course credit or a fixed payment, were healthy right-handed volunteers with no history of neurological or psychiatric problems. Informed consent was obtained and the study was approved by the Internal Review Board at the Leiden University Medical Centre. Two subtests of the Wechsler adult intelligence scale (WAIS-III; Wechsler, 2000), Similarities and block design, were administered to obtain estimates of the participants’ IQs. Performance on these tasks resulted in estimated IQs between 87 and 135, with an average of 109, confirming that participants had average or above average intelligence scores.

2.2. Experimental task

The task was a rule-learning task based on prior research (Crone et al., 2004; Crone, Somsen, Zanolie, & van der Molen, 2006). Two displays, a stimulus display and an outcome display (see Fig. 1a), were presented on each trial. Participants were told to help a dog find its way home by pressing one of four keys corresponding to four doors from left to right. The keys were mapped onto the middle and index fingers of both hands.

Each trial was initiated by a 500 ms fixation mark, followed by 2500 ms of stimulus presentation. The stimulus consisted of a display of two houses, each containing two doors, with the fixation mark between the two houses (Fig. 1).

The dog was displayed randomly below one of the four doors. Participants were instructed to press one of four keys based on three S–R mapping rules they had learned during pre-scan behavioral training (see Section 2.3). They had to respond within the 2500 ms timeframe during which the stimulus was presented; otherwise the feedback “Te langzaam” (too slow) was presented. The trials for which responses were too slow were not included in the analysis. The feedback display was presented for 1000 ms. When the participant responded within the 2500 ms timeframe, the feedback display consisted of the same two houses with four doors and a fixation mark, with an additional ‘+’ or ‘-’ sign placed in the door selected by the participant during the response time. The ‘+’ sign represented a positive feedback, and a ‘-’ sign represented a negative feedback. After the feedback an inter-trial interval jitter varying from 2000 to 8000 ms was presented in 25% of trials.

2.3. Rules

Participants had to use three rules to perform the task correctly. Rule A held that the stimulus (the dog) that appeared below one of the four doors (locations) designated a response with the finger compatible to that door (location). Thus, spatially compatible button-presses were required to the location of the stimulus (see Fig. 1b, left). Rule B held that stimuli that appeared in one of the four locations designated a response with the other finger of the same hand (see Fig. 1b, middle). For rule C, stimuli that appeared at any of the four stimulus locations designated a response with the finger assigned to the location two locations away from the stimulus location (see Fig. 1b, right).

Participants were instructed to help the dog find its way home by using one of these three rules. They were told to use the feedback, presented in the chosen door, to find the appropriate rule for each trial. The participants were told that the dog’s house could change from time to time without any advance warning and that they had to use the trial-by-trial feedback to find its new location. The

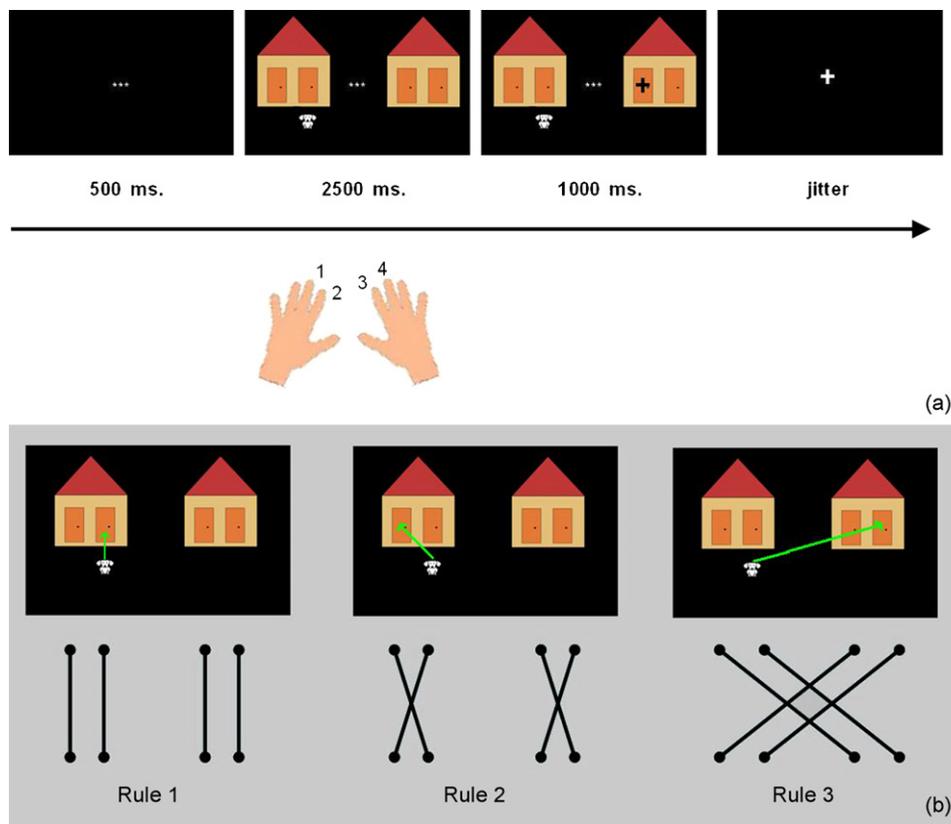


Fig. 1. (a) During scanning, participants viewed a 500 ms fixation cross, followed by a 2500 ms stimulus display. The stimulus required a left index, left middle, right middle or right index finger response, according to three rules that were learned prior to scanning. The stimulus display was followed by a feedback display which was presented for 1000 ms. Random jitter of 2000, 4000, 6000 or 8000 ms was added to all trials. (b) Response-rules: participants were instructed to respond to the stimuli according to three rules. See text for explanation.

rule switched unexpectedly after applying the rule correctly for two, three or four consecutive trials, without any switch-cue given to the participant.

2.4. Design and procedure

Prior to scanning, participants were familiarized with the task during a practice session outside of the scanner. They first learned the three rules described above and their stimulus–response associations. Each rule was learned consecutively in separate practice blocks, during which it remained constant, with no switching. A rule had to be applied correctly on 80% of the trials, with a minimum of 10 trials. When this criterion was met the next rule was learned until all three rules were learned.

After practicing the three rules separately, participants practiced the task as it would be presented in the scanner. This practice block consisted of 100 trials in which the three rules were intermixed. Three such blocks, each lasting approximately 8 min, were presented to participants in the scanner. During the scanning session, participants responded with two response boxes strapped onto their legs. Two buttons were used on each response box in order to keep the S–R mapping equal to the practice block.

The two WAIS subtests ‘similarities’ and ‘block design’ were given following the completion of the scanning session and took approximately 30 min to administer. Including instructions and breaks, participants spent approximately 2 h in the laboratory.

2.5. Scoring methods

After completion of the task the trials were scored as a function of five different types of feedback. Our scoring method benefited from a study by Barcelo and Knight (2002) in which errors were scored as a function of past contextual information (see Fig. 2 for an example of the different types of feedback). We defined (1) ‘first-warning’ negFB as the first-negative feedback given to a participant after a successfully completed sequence of rule-application. This negative feedback signaled a rule switch by indicating that the previously applied sorting rule was no longer correct. An (2) ‘efficient’ negFB indicated that the rule chosen when searching for the appropriate rule was incorrect (i.e. hypothesis testing). For example, if the participant was applying rule A and received a first-warning feedback, either rule B or C should be applied during the next trial. The participant had a 50% chance of applying the correct rule after the rule-switch. When the incorrect rule was chosen first, e.g. rule B, the participant then knew for certain which rule to apply on the next trial (rule C). In this way the ‘efficient’ negFB gave the participant certain knowledge of the appropriate sorting-rule. (3) ‘Error-related’ negFB trials consisted of those trials where the participants failed to apply the correct response in a series of correct trials.

These three negative feedback types were contrasted with two positive feedback (posFB) types. The first type of positive feedback was the (4) ‘first’ posFB, which indicated that the correct rule was found immediately following a rule-switch. (5) ‘Correct’ posFB indicated correct rule use. For an example of the different types of feedback see Fig. 2.

2.6. Data acquisition

Scanning was performed with a standard whole-head coil on a 3.0 T Philips Achieva MR scanner at the Leiden University Medical Centre. Functional data

were acquired using T2*-weighted echo planar imaging (EPI) during three functional runs of 232 volumes each, of which the first two were discarded to allow for equilibration of T1 saturation effects (TR = 2.211 s, TE = 30 ms, ascending interleaved acquisition, 38 slices of 2.75 mm, field of view 220 mm, 80 × 80 matrix, in-plane resolution 2.75 mm). High-resolution T1-weighted anatomical images were also collected after the functional runs. Head motion was restricted using a pillow and foam inserts that surrounded the head. Visual stimuli were projected onto a screen that was viewed through a mirror.

2.7. 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. Structural and functional volumes were spatially normalized to T1 and EPI templates, respectively. The normalization algorithm used a 12-parameter affine transform 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, Kollokian, Kwan, & Evans, 1997), an approximation of Talairach space (Talairach & Tournoux, 1988). Functional volumes were spatially smoothed with an 8 mm FWHM isotropic Gaussian kernel. Statistical analyses were performed on individual participants’ 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 was modeled as an event of interest. 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 pair-wise 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 participants 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 $p < 0.001$. Region-of-interest (RoI) analyses were performed to further characterize rule sensitivity of four a priori predicted regions (DLPFC, medial PFC/ACC, lat-OFC, caudate). RoI analyses were performed with the Marsbar toolbox in SPM2 (Brett, Anton, Valabregue, & Poline, 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. The contrast used to generate functional RoIs was based on an *F*-test for all positive versus all negative feedback stimuli. For all RoI analyses, effects were considered significant at an α of 0.01 (corrected for multiple comparisons).

3. Results

3.1. Behavioral data

Out of 300 trials there were an average of 58.9 (S.D. = 5.29) first-warning error trials, for which the negative feedback signaled a sorting rule change. On approximately half of the rule shifts, participants first made an efficient error in order

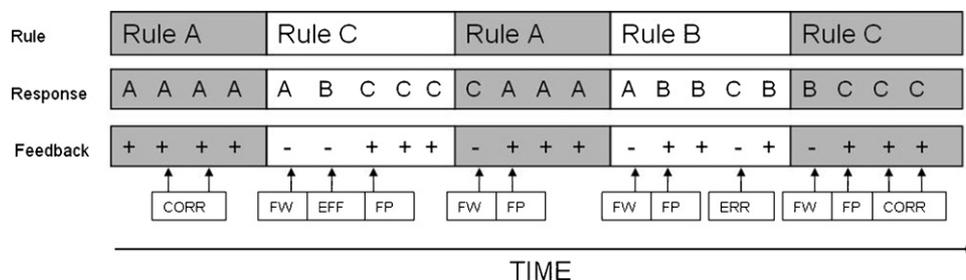


Fig. 2. Illustration of the different types of feedback: FW, first-warning negative feedback; EFF, efficient negative feedback; ERR, error negative feedback; FP, first-positive feedback; CORR, correct positive feedback. See text for explanation of scoring method.

to find the correct sorting rule ($M=26.5$, $S.D. = 5.29$). Finally, on approximately half of the rule shifts participants committed a performance error when applying the correct rule ($M=20.64$, $S.D. = 14.53$). The difference between the number of efficient errors and performance errors was not significant ($F(1, 19) = 1.83$, $p = .19$).

3.2. *Imaging data*

3.2.1. *RoI analysis*

Region-of-interest (RoI) analyses were performed for medial PFC/ACC, right DLPFC, and right lat-OFC based on an F -contrast for negative versus positive feedback, and based on a

sphere for the caudate. For each RoI, activation levels relative to a fixation baseline for each condition and participant were extracted and submitted to repeated measures ANOVAs.

3.2.1.1. *Medial PFC/ACC (BA 32)*. Neural activity in left ACC differed depending on the informative value of the feedback, $F(4, 76) = 20.97$, $p < .001$. Fig. 3 demonstrates that for left ACC, more activation was seen following error-related negFB than following first-warning negFB, $t(19) = 4.08$, $p < .001$. The increase in neural activation in left ACC was also larger for error-related negFB than for efficient negFB, $t(19) = 6.37$, $p < .001$ and for first-warning negFB than for efficient negFB, $t(19) = 3.19$, $p < .005$. In contrast, neural activation following efficient negFB

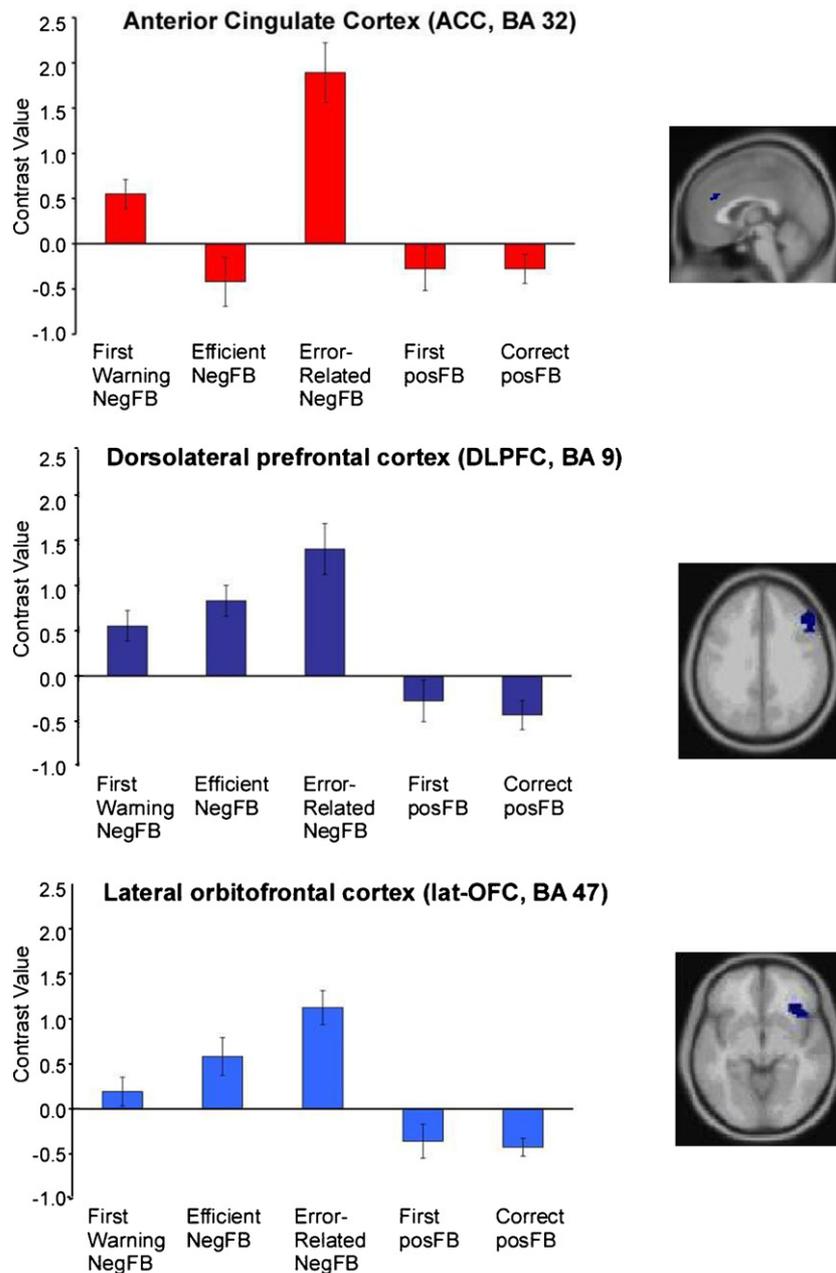


Fig. 3. Activation profiles for RoIs identified on the basis of the F -contrast negative feedback vs. positive feedback ($p < .001$). These RoIs included right DLPFC (BA 9, $x = 44$, $y = 32$, $z = 32$), medial PFC/ACC (BA 32, $x = -1$, $y = 35$, $z = 27$) and right lat-OFC (BA 47/13, $x = 36$, $y = 24$, $z = -8$). Error bars depict an estimate of within-subject standard error. * $p < .05$.

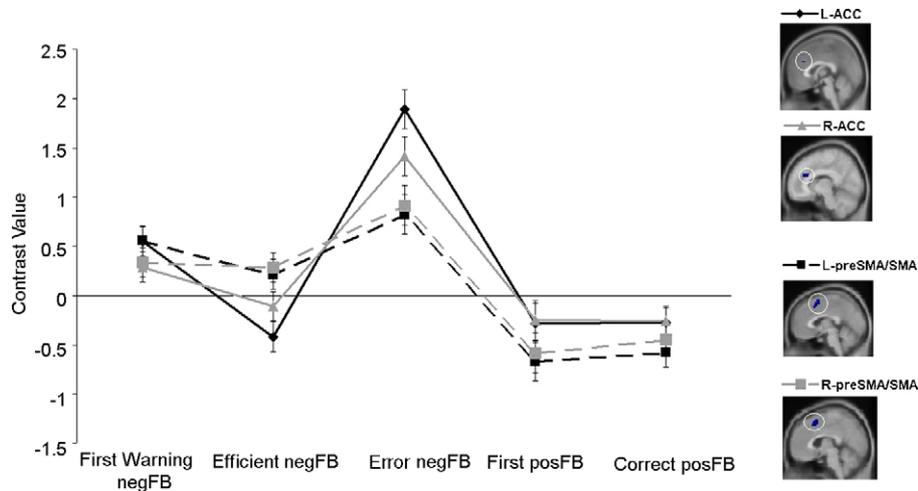


Fig. 4. Activation profiles for ACC and pre-SMA RoIs identified on the basis of the F -contrast negative feedback vs. positive feedback ($p < .001$). These RoIs included left ACC (BA 32, $x=0$, $y=33$, $z=29$), right ACC (BA 32, $x=10$, $y=30$, $z=26$), left pre-SMA (BA 6, $x=-2$, $y=13$, $z=52$) and right pre-SMA (BA 6/8, $x=4$, $y=16$, $z=51$). Error bars depict an estimate of within-subject standard error.

did not differ from neural activation following first posFB or correct posFB, $t(19) = .57$, $p = .58$ and $t(19) = .55$, $p = .58$, respectively. There was no significant difference in activation for first posFB or correct posFB, $t(19) = .01$, $p = .99$.

Prior research has remained inconclusive about the role of medial PFC subregions in error and feedback processing. We separated the functional RoI for the broader medial PFC region in left and right ACC (BA 32) and left and right (pre-supplementary) motor area (pre-SMA/SMA) (BA 6/8). As can be seen in Fig. 4, the pattern of activation was similar for ACC and pre-SMA/SMA, but the strength of the effects was stronger for ACC than pre-SMA/SMA (region \times feedback interaction, $F(4, 76) = 12.86$, $p < .001$), especially for left ACC (region \times feedback \times hemisphere, $F(4, 76) = 5.72$, $p < .001$). Post-hoc region \times hemisphere comparisons for each feedback type resulted in significant region effects for efficient negFB, with higher activation in pre-SMA/SMA than in ACC ($F(1, 19) = 10.04$, $p < .005$), and error negFB ($F(1, 19) = 10.36$, $p < .001$), with higher activation in ACC than in pre-SMA/SMA. The latter effect was modulated by an effect of hemisphere (region \times hemisphere, $F(1, 19) = 8.96$, $p < .01$). Activation following error negFB was more pronounced in both left and right ACC relative to left and right pre-SMA (p 's $< .05$), but the difference was larger for left ACC than for right ACC ($p < .05$). Significant region effects were also found for first posFB ($F(1, 19) = 6.13$, $p < .05$) and correct posFB ($F(1, 19) = 5.16$, $p < .05$), with higher activation for ACC than for pre-SMA. These results strengthen the conclusion that feedback sensitivity is most pronounced in ACC.

3.2.1.2. DLPFC (BA 9). Activation in DLPFC also differentiated among the five possible types of feedback, $F(4, 76) = 20.34$, $p < .001$. As shown in Fig. 3, error-related negFB elicited more activation than first-warning negFB, $t(19) = 3.69$, $p < .005$, whereas neural activity for efficient negFB and error-related negFB did not differ, $t(19) = 1.77$, $p = .093$. These results indicate that DLPFC was least active following first-warning negFB

and showed increased activation following both efficient negFB and error-related negFB. All negative feedback types resulted in more activation in DLPFC relative to positive feedback types (all p 's $< .05$). Neural activity for first posFB and correct posFB did not differ from each other, $t(19) = 0.93$, $p = .362$. Right DLPFC did not differ in activation profile from left DLPFC (region \times feedback, $F < 1$).

3.2.1.3. DLPFC versus ACC. The next comparison tested whether ACC and DLPFC were differentially active following first-warning negFB and efficient negFB. We predicted a two-way interaction between region (ACC versus DLPFC) \times feedback (first-warning negFB, efficient negFB, error negFB); as expected, this interaction was highly significant, $F(2, 38) = 12.30$, $p < .001$. Follow-up comparisons indicated that ACC and DLPFC did not differ in activation following first-warning negFB, $F(1, 19) = .01$, $p = .995$, and error-related negFB, $F(1, 19) = 2.40$, $p = .14$. In contrast, DLPFC was more active than ACC following efficient negFB, $F(1, 19) = 23.77$, $p < .001$. These findings confirm the hypothesis that ACC and DLPFC are differentially sensitive to the informative value of negative feedback.

Next, we tested for differences in neural activity associated with first posFB versus correct posFB with a region (ACC versus DLPFC) \times feedback (first positive versus correct) ANOVA. This analysis did not result in significant effects across regions (main effect, $F(1, 19) = .22$, $p = .64$), or between regions (region \times feedback, $F(1, 19) = .89$, $p = .36$).

3.2.1.4. Lateral OFC (BA 47). Lateral OFC was also differentially sensitive to the informative value of the feedback, $F(4, 76) = 17.48$, $p < .001$. A region (DLPFC versus lat-OFC) \times feedback (5) ANOVA indicated that the pattern of feedback-related activation did not differ between lat-OFC and DLPFC, $F(4, 76) = .97$, $p = .42$. In contrast, a region (lat-OFC versus medial PFC/ACC) \times feedback (5) ANOVA showed that the pattern of activity associated with feedback types was dif-

ferent for lat-OFC relative to ACC, $F(4, 76) = 12.28, p < .001$. Separate analyses for lat-OFC confirmed the similarity in activation as seen for DLPFC. These results show more activation after error-related negFB than after a first-warning negFB, $t(19) = 4.01, p = .001$. First-warning negFB and efficient negFB showed no significant difference in activation in the lat-OFC, $t(19) = 1.52, p = .132$. Finally, as with DLPFC, there was no difference in activation for efficient negFB and error-related negFB, $t(19) = 1.64, p = .118$ or first posFB and correct posFB, $t(19) = 0.57, p = .579$. Right lat-OFC did not differ in activation profile from left lat-OFC (region \times feedback, $F < 1$).

3.2.1.5. Caudate. In addition to medial PFC/ACC, DLPFC and lat-OFC, the caudate has been implicated in the processing of feedback (Nieuwenhuis et al., 2005; Tricomi et al., 2006; Ullsperger & von Cramon, 2003; van Veen et al., 2004). However, in this study neither left nor right caudate showed significant differences in neural activation for negative versus positive feedback or vice versa. No ROIs based on feedback (neg > pos or pos > neg) or general (all > fixation) contrasts could be selected. Additional analysis of spherical ROIs with radii 5 and 10 mm, centered on coordinates $\pm 8, 8, 8$ (Tricomi et al., 2006), did not result in any significant effect of feedback type.

3.2.2. Additional analysis: first-positive feedback

An additional analysis was performed to tease apart any potential differences in activation due to different meanings of the first posFB. We tested for differences in neural activity associated with a first posFB that followed a first-warning negFB, an efficient negFB and error-related negFB, because these can have different informative values. For ACC, the preceding trial type did not result in different neural activity, $F(2, 38) = 0.02, p = .98$. In contrast, first posFB activation in DLPFC depended on the preceding trial type, $F(2, 38) = 5.34, p < .01$. Post-hoc analysis showed that first posFB after error-related negFB resulted in greater activation than first posFB following first-warning negFB, $t(19) = 2.37, p < .05$, or efficient negFB, $t(19) = 2.77,$

$p < .01$, whereas the latter two did not differ from each other, $t(19) = 0.34, p = .73$.

3.2.3. Additional analysis: first-warning errors

A second additional analysis was performed in order to investigate whether the participant expected first-warning negFB following a certain number of correct consecutive trials. A random switch in the rule type could follow after correctly applying a rule for two, three or four consecutive trials. It is possible that the participant learned during the task that four consecutive correct trials were always followed by first-warning negFB on the fifth rule application. In order to investigate whether there were differences in activity for a first-warning negFB after two, three, or four correct trials, three types of first-warning negFB were classified. A repeated measures ANOVA showed that there was no effect of first-warning negFB type for DLPFC, $F(2, 38) = 0.43, p = .52$, or ACC, $F(2, 38) = 2.00, p = .15$. These results suggest that participants were not aware that after four correct applications of the correct rule a rule-switch would occur.

3.2.4. Whole-brain analysis

In addition to the ROI analysis a whole-brain analysis was conducted, using a p -value threshold of .001, uncorrected for multiple comparisons. As expected, a comparison for negative > positive feedback resulted in increased activation in the medial PFC, including ACC (BA 32) and pre-SMA/SMA (BA 6/8), lat-OFC (BA 47) and DLPFC (BA 9); consistent with prior studies, activation was most pronounced in the right hemisphere (Konishi et al., 2002). See Fig. 5 for an illustration and Table 1 for the coordinates.

4. Discussion

The aim of this study was to investigate the neural correlates of processing performance feedback with different informative values. Medial PFC/ACC, DLPFC, and lat-OFC were fractionated in order to test their differential sensitivity to the informative

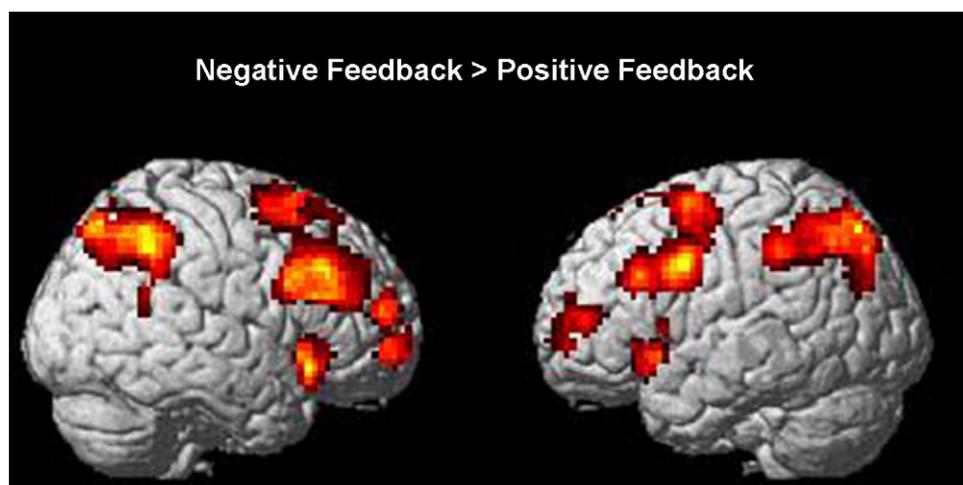


Fig. 5. Whole-brain activation associated with negative feedback processing (negative feedback > positive feedback). The presentation of negative feedback resulted in activation in bilateral DLPFC, lat-OFC, superior and inferior parietal cortex and medial PFC (ACC and pre-SMA/SMA). The coordinates for these regions are presented in Table 1.

Table 1

Regions implicated in feedback processing, for the contrast negative feedback > positive feedback processing contrast, $p < .001$, at least 10 contiguous voxels

Region of activation	~BA	x	y	z
R medial frontal gyrus	8/32/24	9	27	39
L medial frontal gyrus	6/32/24	-6	15	45
R middle frontal gyrus	9	51	30	33
L middle frontal gyrus	10	-36	45	12
R inferior frontal gyrus	47	36	21	-12
L inferior frontal gyrus	47	-33	27	-6
L precentral gyrus	6	-48	3	36
R superior frontal gyrus	10	33	60	-9
L superior frontal gyrus	10	-27	57	6
L insula	13	-48	12	3
L superior parietal cortex	7	-30	-69	45
L inferior parietal cortex	40	-45	-39	39

value of negative and positive feedback. In prior studies, it had been shown that these regions play a role in feedback processing (Konishi et al., 2002; Lie et al., 2006; Monchi et al., 2001), but their contributions to different types of feedback remained unclear. In part, this is because the paradigms used in the previous work featured task design or ambiguous trials that did not allow differentiating the informative value of the feedback (e.g. Holroyd et al., 2004; Nieuwenhuis et al., 2005; van Veen et al., 2004). Our aim in the current study was to distinguish different feedback types by means of a rule-switch task without ambiguous trials (see Walton et al., 2004), allowing us to tease apart the different information values of negative and positive performance feedback (see also Hampshire & Owen, 2006, for a similar approach).

The behavioral data indicated that participants made efficient use of performance feedback for rule-switching. Because participants were switching between three types of rules, the finding that participants made efficient errors on approximately 50% of the trials shows that participants were successfully searching and switching between rules (“if the correct rule is no longer A, then it must be rule B or C”). Participants also made a considerable number of performance errors during rule application, which is consistent with prior results using the same task in adults (Crone et al., 2004). The main question for the imaging results was whether these different types of errors and feedback were neurally dissociable.

As expected, the imaging results provide evidence for such a dissociation in DLPFC and medial PFC/ACC. These regions were differently active depending on the informative value of negative feedback. In particular, DLPFC was more active than medial PFC/ACC following efficient negFB, which signaled that possibilities for future actions were constrained and required goal-directed behavior. In contrast, medial PFC/ACC was relatively more active following first-warning negFB, which signaled that prior performance was no longer correct, relative to efficient negFB. Both regions were equally sensitive to feedback indicating a performance error.

Previous studies have suggested that DLPFC is generally involved in monitoring of task sets in working memory (Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2005; Kerns

et al., 2004; Monchi et al., 2001). According to Petrides (2000), DLPFC subserves complex operations such as monitoring and manipulation of working memory contents. It is therefore not surprising that DLPFC is recruited when an efficient error is made. An efficient error indicates that the current choice made while searching for the appropriate rule is incorrect, but the participant knows with certainty which rule is correct on the next trial and can employ goal-directed behavior. In order to accomplish this, the participant is required to update and check information regarding which rules are incorrect in working memory to determine the correct rule. Thus, greater DLPFC activation after making an efficient error than after a first-warning error indicates that this region does indeed contribute to the use of goal-directed behavior to apply the correct rule on the next trial (Walton et al., 2004). In general, lateral PFC may be more sensitive to the informative value of the received feedback than to the valence of the feedback per se.

Medial PFC/ACC showed a reversed pattern of neural activation compared to DLPFC. The results demonstrate that medial PFC/ACC, in particular ACC, was more active when a first-warning error was committed than when an efficient error was committed. These findings suggest that medial PFC/ACC is more involved in processes concerned with violation of expectations. When a first-warning error has been made, the participant has to engage in hypothesis testing to find the new correct rule but does not yet know which behavior will be correct on subsequent trials. Prior studies have shown that the ACC is important when an action has not produced the desired result, or when there is a violation of reward prediction (Mars et al., 2005). In particular, ACC is thought to indicate that ongoing events are unexpectedly disadvantageous and that this information can subsequently be used to guide adaptive behavior (Holroyd & Coles, 2002). The finding that medial PFC/ACC was most active following first-warning negative feedback and following feedback indicating that an error was committed supports this theory (see also Ullsperger & von Cramon, 2001). Importantly, following efficient negative feedback, medial PFC/ACC was not more active than following positive feedback, suggesting that medial PFC/ACC is not sensitive to negative feedback per se. The finding that medial PFC/ACC is differentially sensitive to the type of negative feedback may explain previous findings that failed to show an effect of processing negative and positive in medial PFC (Nieuwenhuis et al., 2005; van Veen et al., 2004). Together, the results of this study indicate that medial PFC/ACC is not generally involved in the processing of negative feedback but is specifically sensitive to feedback signaling a violation of expectations. Further, it is unlikely that activation in ACC is the result of differences in the frequency with which negative feedback is presented (Nieuwenhuis et al., 2005), because first-warning negative feedback occurred twice as often as efficient negative feedback but still elicited more activation in medial PFC/ACC.

In prior studies, the orbitofrontal section of the PFC (OFC) has been implicated in reward-related processing of information (Elliott, Friston, & Dolan, 2000; O’Doherty et al., 2003). Although a number of important distinctions has been proposed between the functions of DLPFC and OFC (e.g. McClure,

Laibson, Loewenstein, & Cohen, 2004), the contribution of each of these regions to feedback processing has remained unclear. As predicted from previous studies, the lateral portion of the OFC was more active following negative feedback than following positive feedback (Elliott et al., 2000; O'Doherty et al., 2003), and activation in this region was dependent on the type of negative feedback. However, unlike other studies, the region did not differ in activation pattern from DLPFC (Hampshire & Owen, 2006), suggesting that lat-OFC may have a more specific role in feedback processing rather than simply signaling the valence of the feedback (O'Doherty et al., 2003). Contrary to expectations, the caudate was not sensitive to any feedback type in this study. This was surprising because prior research has proposed that the caudate works closely together with prefrontal cortex regions in feedback processing and set-shifting (Monchi et al., 2006). However, it has also been suggested that the caudate is sensitive to learning signals in general, which could explain why in some studies this region is sensitive to positive feedback (Tricomi et al., 2006) and in others to negative feedback (Monchi et al., 2006). The specific role of these regions should be examined in future research.

In conclusion, the different patterns of activation for medial PFC/ACC and DLPFC support the hypothesis that these regions are differentially sensitive to the informative value of performance feedback, and disconfirm the hypothesis that these regions are only sensitive to the valence of the feedback. In particular, medial PFC/ACC is most active in situations in which expectations are violated and where there are multiple response options. In contrast, DLPFC is most active in situations where goal-directed behavior is needed due to restricted response alternatives. These results are consistent with models showing that medial PFC/ACC signals violations of reward prediction (Holroyd & Coles, 2002) and DLPFC is important for goal-directed behavior (Miller & Cohen, 2001). A recent study used resting-state analysis to demonstrate that DLPFC and ACC are part of different networks (Dosenbach et al., 2007), each of which aid in the adjustment of behavior in separate ways. In future studies, it will be important to relate resting-state (default) networks to task performance.

In prior research using heart rate indices, we investigated developmental changes in rule-shifting and feedback processing in participants aged 8–25. We demonstrated that the processing of first-warning feedback reached mature levels earlier than the processing of performance-errors, providing further evidence for process dissociability in feedback processing. We are currently testing the effects that informative value of performance feedback has on activation patterns in medial PFC/ACC and DLPFC in children aged 8–18 using fMRI. While it is known that the prefrontal areas of the brain are maturing until adulthood both structurally (Sowell et al., 2004) and functionally (Crone et al., 2006; Diamond, 2002; Durston & Casey, 2006), it is interesting to see whether regions within PFC follow different developmental trajectories, which may contribute to our understanding of developmental differences in goal-directed behavior and hypothesis testing when switching between rules.

Acknowledgements

The third and last authors (SARBR and EAC) were supported by NWO-VENI/VIDI grants. We thank Santani Teng and Michiel Westenberg for helpful comments on an earlier version of the manuscript, and Wouter Teeuwisse for technical assistance.

References

- Barcelo, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal changes. *Neuropsychologia*, *40*, 349–356.
- Bayless, S. J., Gaetz, W. C., Cheyne, D. O., & Taylor, M. J. (2006). Spatiotemporal analysis of feedback processing during a card sorting task using spatially filtered MEG. *Neuroscience Letters*, doi:10.1016/j.neulet.2006.09.073.
- Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). *Region of interest analysis using an spm toolbox*. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Cocosco, C. A., Kollokian, V., Kwan, R. K. S., & Evans, A. C. (1997). Brain web: Online interface to a 3D MRI simulated brain database. *NeuroImage*, *5*, S425.
- Crone, E. A., Ridderinkhof, K. R., Worm, M., Somsen, R. J. M., & van der Molen, M. W. (2004). *Developmental Science*, *7*, 443–455.
- Crone, E. A., Somsen, R. J., Zanolie, K., & van der Molen, M. W. (2006). A heart rate analysis of developmental change in feedback processing and rule shifting from childhood to early adulthood. *Journal of Experimental Child Psychology*, *95*(2), 99–116.
- Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L., & Bunge, S. A. (2005). Neurocognitive development of the ability to manipulate information in working memory. In *Proceedings of the National Academy of Sciences of the United States of America*, *103* (pp. 9315–9320).
- Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin card sorting test to frontal and lateralized frontal brain damage. *Neuropsychology*, *17*, 255–264.
- Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Science of the United States of America*, *104*(26), 11073–11078.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In S. A. Knight (Ed.), *The frontal lobes*. London: Oxford University Press.
- Dias, R., Robbins, T. W., & Roberts, A. C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analogue of the Wisconsin card sorting test: Restriction to novel situations and independence from “on-line” processing. *Journal of Neuroscience*, *17*(23), 9285–9297.
- Dove, A., Pollmann, S., Schubert, T., Wiggins, C. J., & Von Cramon, D. Y. (2000). *Cognitive Brain Research*, *9*, 103–109.
- Durston, S., & Casey, B. J. (2006). A shift from diffuse to focal cortical activity with development: The author's reply. *Developmental Science*, *9*, 18–20.
- Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience*, *20*, 6159–6165.
- Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, *3*, 516–520.
- Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*, 2279–2282.
- Hampshire, A., & Owen, A. M. (2006). Fractionating attentional control using event-related fMRI. *Cerebral Cortex*, *16*, 1679–1689.
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.
- Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., et al. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*, 497–498.
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.

- Konishi, S., Hayashi, T., Kikyo, H., Takahashi, E., & Miyashita, Y. (2002). Hemispheric asymmetry in human lateral prefrontal cortex during cognitive set-shifting. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(11), 7803–7808.
- Lie, C., Specht, K., Marshall, J. C., & Fink, G. R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin card sorting task. *NeuroImage*, *30*, 1038–1049.
- Mars, R. B., Coles, M. G. H., Grol, M. J., Holroyd, C. B., Nieuwenhuis, S., Hulstijn, W., et al. (2005). Neural dynamics of error processing in medial frontal cortex. *NeuroImage*, *28*, 1007–1013.
- McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex functioning. *Annual Review of Neuroscience*, *24*, 167–202.
- Miltner, B. (1963). Effects of different brain regions on card sorting. *Archives of Neurology*, *9*, 100–110.
- 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. *Journal of Cognitive Neuroscience*, *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. *Journal of Neuroscience*, *21*(19), 7733–7741.
- Monchi, O., Petrides, M., Strafella, A. P., Worsley, K. J., & Doyon, J. (2006). Functional role of the basal ganglia in the planning and execution of actions. *Annals of Neurology*, *59*, 257–264.
- Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience and Biobehavioral Reviews*, *28*, 441–448.
- Nieuwenhuis, S., Slagter, H. A., Alting von Geusau, N. J., Heslenfeld, D. J., & Holroyd, C. B. (2005). Knowing good from bad: Differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience*, *21*, 3161–3168.
- O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *The Journal of Neuroscience*, *23*, 7931–7939.
- Petrides, M. (2000). The role of the mid-dorsolateral prefrontal cortex in working memory. *Experimental Brain Research*, *133*, 44–54.
- Rushworth, M. F., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Science*, *8*, 410–417.
- Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *The Journal of Neuroscience*, *24*, 8223–8231.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Tricomi, E., Delgado, M. R., McClelland, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience*, *18*(6), 1029–1043.
- Ullsperger, M., & von Cramon, D. Y. (2001). Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *NeuroImage*, *14*, 1387–1401.
- 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. *Journal of Neuroscience*, *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 & Cognition*, *56*, 267–276.
- Walton, M. E., Devlin, J. T., & Rushworth, M. F. S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience*, *7*, 1259–1265.
- Wechsler, D. (2000). *WAIS-III: Nederlandse bewerking Wechsler adult intelligence scale—Derde Editie*. Lisse: Swets & Zeitlinger B.V.