Errors without conflict: Implications for performance monitoring theories of anterior cingulate cortex

Vincent van Veen\textsuperscript{a}, Clay B. Holroyd\textsuperscript{b}, Jonathan D. Cohen\textsuperscript{b,c}, V. Andrew Stenger\textsuperscript{d}, Cameron S. Carter\textsuperscript{e,*}

\textsuperscript{a} Department of Psychology, University of Pittsburgh, PA, USA
\textsuperscript{b} Department of Psychology, Princeton University, NJ, USA
\textsuperscript{c} Department of Psychiatry, University of Pittsburgh, PA, USA
\textsuperscript{d} Department of Radiology, University of Pittsburgh, PA, USA
\textsuperscript{e} Departments of Psychiatry and Psychology, University of California at Davis, 4701 X Street, Sacramento, CA 95817, USA

Accepted 16 June 2004
Available online 3 September 2004

Abstract

Recent theories of the neural basis of performance monitoring have emphasized a central role for the anterior cingulate cortex (ACC). Replicating an earlier event-related potential (ERP) study, which showed an error feedback negativity that was modeled as having an ACC generator, we used event-related fMRI to investigate whether the ACC would differentiate between correct and incorrect feedback stimuli in a time estimation task. The design controlled for response conflict and frequency and expectancy effects. Although participants in the current study adjusted their performance following error feedback, we did not observe error feedback-evoked ACC activity. In contrast, we did observe ACC activity while the same subjects performed the Stroop task, in which an area of the ACC activated during both conflict and error trials. These findings are inconsistent with previous dipole models of the error feedback negativity, and suggest the ACC may not be involved in the generation of this ERP component. These results question involvement of the ACC in the detection of errors per se when controlling for conflict.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Functional MRI; Anterior cingulate cortex; Frontal lobes; Errors; Feedback; Attention; Performance monitoring; Executive control; Stroop interference

1. Introduction

The anterior cingulate cortex (ACC) is usually thought of as playing an important role in implementing the processes underlying adjustments of performance control. In particular, it is thought to contribute to the processing of errors.

Several lines of research have linked ACC activity to the presence of “slips” (fast and impulsive but erroneous responses) in speeded response tasks. Event-related potential (ERP) research has shown the presence of a frontocentral negativity immediately following an action slip (Coles, Scheffers, & Fournier, 1995; Coles, Scheffers, & Holroyd, 2001; Falkensteine, Hoormann, Christ, & Holhsbein, 2000; Gehring, Goss, Coles, Meyer, & Donchin, 1993; Scheffers, Coles, Bernstein, Gehring, & Donchin, 1996). This component, referred to as error negativity or error-related negativity (ERN), has consistently been modeled as being generated by the ACC (Alain, McNeely, He, Christensen, & West, 2002; Dehaene, Posner, & Tucker, 1994; Holroyd, Dien, & Coles, 1998; Luu, Tucker, Derryberry, Reed, & Poulsen, 2003; Mathalon, Whitfield, & Ford, 2003; Van Veen & Carter, 2002b). In support of these models, neuroimaging studies using fMRI have shown increased ACC activity...
during error trials (Carter et al., 1998; Carter, MacDonald, Ross, & Stenger, 2001; Kerns et al., 2004; Kiehl, Liddle, & Hopfinger, 2000; Menon, Adleman, White, Glover, & Reiss, 2001). Theories differ, however, on the cognitive processes that it implements—that is, theories differ on the nature of the algorithms that are carried out by the ACC, and also on the nature of the representations that these algorithms work on. Two recent theories can be distinguished.

According to one theory, by Holroyd and Coles (2002), the ERN is generated as part of a reinforcement learning process. According to this “reinforcement learning” (RL) theory, behavior is monitored by an “adaptive critic” in the basal ganglia, that detects whether events are either better or worse than expected, and signals this with a phasic increase or a decrease, respectively, in dopaminergic activity. According to this proposal, the function of the ACC is to select between mental processes competing for access to the motor system, and the ERN is assumed to be generated because the inhibitory influence of the dopaminergic innervation in the ACC is briefly disrupted, fine-tuning the ACC to do a better job on future trials.

Another view holds that the ACC monitors for the presence of conflict between simultaneously active but incompatible processing streams (for reviews, see Botvinick, Braver, Barch, Carter, & Cohen, 2001; Carter, Botvinick, & Cohen, 1999; Cohen, Botvinick, & Carter, 2000; Gruber & Goschke, 2004; Van Veen & Carter, 2002a). Van Veen and Carter (2002b) have suggested that during errors, conflict is considered to arise between the executed incorrect response and activation of the correct response due to ongoing stimulus evaluation. This theory, which can account for many observations of ACC activation during both correct trials and tasks that elicit response conflict and during errors in functional MRI and ERP research, has also received extensive support from the results of computational modeling studies (Botvinick et al., 2001; Jones, Cho, Nystrom, Cohen, & Braver, 2003; Yeung, Botvinick, & Cohen, in press). In these modeling studies, conflict is operationalized as Hopfield’s energy (Hopfield, 1982), which is a measure of activity taken over all units in the set of interest and how strongly these units are compatible with one another; a high-energy state might be the result of crosstalk between different streams of information processing. This theory sees the ACC as detecting such high-energy states, and engaging areas involved in top-down control which act to lower the energy state. It is important to note that, while the ACC appears to be engaged mostly by conflict between response representations (Milham et al., 2001; Van Veen, Cohen, Botvinick, Stenger, & Carter, 2001), and in the modeling studies so far energy is only computed over the response layer (Botvinick et al., 2001; Jones et al., 2003), it has always been considered that other forms of representa-

tional conflict may also be monitored by the ACC (Botvinick et al., 2001; Carter et al., 1998; Carter et al., 1999). Recently, reports have suggested that the ACC may indeed be engaged by conflict at other levels of processing (Badre & Wagner, 2004; Milham, Banich, & Barad, 2003; Van Veen & Carter, 2002a; Weissman, Giesbrecht, Song, Mangun, & Woldorff, 2003). At the computational level novel or unexpected stimuli, too, may induce a high-energy state (conflict between an incoming and expected stimulus), and such stimuli have also been observed to evoke ACC activity (Clark,annon, Lai, Benson, & Bauer, 2000; Dien, Spencer, & Donchin, 2003; Downar, Crawley, Mikulis, & Davis, 2000; Kiehl, Laurens, Duty, Forster, & Liddle, 2001) although in some cases this conflict can also be understood in terms of conflict between frequent, prepared responses and rare ones.

According to both the RL and conflict theories, ACC activity may be related to competing processes, and both theories appear to make similar predictions regarding slips during performance. However, some differences are readily apparent, in particular regarding its engagement following error feedback. According to conflict theory, ACC activity during error trials is generated immediately during and following the slip and reflected in the ERN, as a result of conflict between the simultaneous activation of the fast erroneous response and the slower buildup of activation of the correct response, as outlined above (Van Veen & Carter, 2002a, 2002b). Conversely, during correct conflict trials, conflict is largest preceding the response and reflected in the frontocentral N2 of the ERP (Van Veen & Carter, 2002a, 2002b). But an error feedback stimulus to which no overt response is required might also evoke conflicts of sorts, to the extent that conflict occurs between an incoming and an expected—or desired—stimulus, or between the active task set and the task set imposed by the feedback stimulus. In situations in which the error feedback stimulus does not elicit such conflict, conflict theory would not predict any ACC activity.

In contrast, according to RL theory, the ACC can also be activated by stimuli that might not evoke conflict. Holroyd and Coles (2002) based their theory partially on evidence from ERPs to feedback stimuli informing subjects about their performance. These ERP waveforms show a stimulus-locked, ERN-like component, which is increased when the stimulus informs the subject that an error has been made (for a review, see Nieuwenhuis, Holroyd, Mol, & Coles, 2004). Source localization has shown that this component can be modeled as being generated by the ACC (Badgaiyan & Posner, 1998; Gehring & Willoughby, 2002; Luu et al., 2003; Milten, Braun, & Coles, 1997; Ruchswor, Grotte, Spitzer, & Kiefer, 2002). According to RL theory, this phenomenon reflects the same basic mechanism as the ERN, namely, a phasic dopamine reduction in
As stated earlier, although the ACC appears to be activated most robustly to conflict between response representations, other results suggest that the ACC can be engaged by conflict between other types of representations. ACC activation has been observed in response to conflict between representations at intermediate (conceptual) levels of processing (Beauregard, Lévesque, & Bourgouin, 2001; Milham et al., 2003; Robertson et al., 2000) or by conflict caused by expectancy violations or novelty (Downar et al., 2000). Thus, a conflict view of ACC functioning would not necessarily predict its activation following error feedback, unless this feedback elicits conflict between conceptual representations, requires a task set reconfiguration, or violates a strong expectancy. The latter might well be the case for ACC activation in the Wisconsin card sorting task (WCST) (Monchi et al., 2001) or other studies in which error feedback is infrequent. Expectancy violations and novelty will most likely involve conflict at intermediate/semantic levels of representation, or at the level of task set implementation. The reinforcement learning theory predicts ACC engagement to error feedback even when such stimuli do not involve novelty or expectancy violation. As such, one way to distinguish between the competing theories is to obtain evidence that the ACC does indeed activate to error feedback, in situations in which a conflict interpretation can be ruled out.

In the study by Miltner et al. (1997), ERPs were measured during a time estimation task in which the participant received feedback which told him or her whether the response was executed correctly or incorrectly. During each trial of this task, participants received a cue, and had to make a response as close as possible to 1 s after the onset of this cue. A variable time window, centered around the moment 1000 ms after cue onset, was used to evaluate this response: if the response fell within this time interval, it was judged to be correct; if the response was made either too fast or too slow, it was judged to be incorrect. The duration of the time window depended on the accuracy of the previous trial: it decreased when the previous response was correct, making it more difficult for the participant, and it increased when the response was incorrect, making it easier. Then, briefly after the response was made, a feedback stimulus informed the participant whether the time estimation was accurate or not. The use of a time window that varied with performance ensured the following two interesting aspects of this task. First, each participant received about as many correct as error feedback stimuli, thus ruling out an interpretation in terms of expectancy violation or novelty. Second, participants were most likely unable to determine or even guess whether a response was correct before they received feedback. Therefore, expectancy-related factors probably did not influence the processing of the feedback stimulus in that study. Thus, in this task, participants engaged in error processing in the absence of conflict processing.

Miltner et al. (1997) analyzed the ERPs locked to the onset of the feedback stimuli, and found that error feedback elicited a negative deflection around 230–270 ms following feedback stimulus onset. Miltner et al. tested participants using visual, auditory and somatosensory feedback stimuli, and for each modality, source localization suggested a generator in the ACC. Similar dipole models of comparable error feedback negativities have also been described by other researchers (Badgaiyan & Posner, 1998; Gehring & Willoughby, 2002; Luu et al., 2003; Ruchsow et al., 2002), although some of these models have implicated more posterior areas of the cingulate cortex in the generation of this component (Badgaiyan & Posner, 1998; Luu et al., 2003).

Some caveats are in order before strong inferences can be made from these ERP studies. Most importantly, source localization is an inherently imprecise method of localizing brain activity. Source localization can only provide simplified models of the brain activity that might have generated a given scalp topography but there is always the chance that a model is not correct, regardless of how well it can explain the data. For any given scalp topography, an infinite possibility of dipole solutions exists. It is therefore impossible to make any definitive claims about brain activity solely on the basis of dipole source models (Peters & de Munck, 1990; Scherg, 1990). Second, the scalp distribution of the feedback negativity in the Miltner et al. (1997) study appears to differ somewhat from that of the ERN, as actually noted by those authors (p. 795). Thus, it is not clear whether the error feedback-related negativity reflects the same cognitive process as the ERN. In fact, as mentioned, a few studies have found that the error feedback-related negativity might be generated by the posterior cingulate rather than the ACC (Badgaiyan & Posner, 1998; Luu et al., 2003).
et al., 2003). Third, Miltnner et al. used only 36 electrodes; whereas it has been argued that many more electrodes are needed to obtain reasonably accurate source models (Srinivasan, Tucker, & Murias, 1998). In fact, both reports that modeled the error feedback-related negativity as being generated by the posterior cingulate cortex (Badgaiyan & Posner, 1998; Luu et al., 2003) used high-density EEG (128 channels plus the average reference channel). In addition, both of those studies compared the error feedback negativity to another ACC-generated component, and both found differences. Luu et al. (2003) compared the error feedback negativity to the ERN, and found that despite the similarity in scalp topography, the ERN was explained by a model that included both anterior and posterior cingulate sources, whereas the error feedback negativity was explained by a posterior cingulate source alone. Badgaiyan and Posner (1998) compared the error feedback negativity to (what appears to be) the frontocentral N2 (thought to be the same component as the ERN, see Van Veen & Carter, 2002b), and found differing scalp topographies and dipole models; the error feedback negativity was explained by a posterior cingulate source, the frontocentral N2 by an anterior cingulate source. Conclusive evidence that the error feedback negativity and the ERN are the same component is therefore lacking; no high density EEG study to date has performed a within-study comparison of the scalp topographies of these two components and found that they are the same. (It should be noted, however, that Luu et al. did not use an exploratory dipole fitting method, leaving open the possibility that an ACC dipole might have provided a better fit to their model.)

For these reasons, an fMRI investigation of brain activity in response to error feedback may shed better light on the underlying generators of this component and the neural substrate of processing error feedback. Therefore, we adapted the paradigm described by Miltnner et al. (1997) for a study using fMRI in order to examine the neural basis of error feedback processing. Because of the slowness of the BOLD response, we increased the stimulus onset asynchronies (SOAs) in the experiment to 12s, allowing enough time for the BOLD response to resolve. Because of this manipulation, we avoided the possibility of cue-related activity confounding the feedback-related activity. In sum, a strong prediction of the reinforcement learning theory is increased activation following error feedback compared to correct feedback, in the absence of conflict related to action planning or expectancy violation. Conflict theory does not predict such activity. In order to compare any possible error feedback-related activity to activity elicited by slips in a speeded response task, participants also performed a version of the Stroop color-word task (MacLeod, 1991; Stroop, 1935).

2. Materials and methods

2.1. Research participants

Informed consent was obtained from 14 healthy young participants (six women, eight men; all right-handed; age range 19–28, \( M = 21.4 \) years, \( SD = 2.2 \)) in agreement with the Institutional Review Board of the University of Pittsburgh. They received $50 per hour for their participation.

2.2. Procedure

Each participant performed four blocks of 16 trials each of the time estimation task. At the beginning of each block, a fixation point (Æ) was shown on screen for 12s. Each trial consisted of a cue stimulus that participants had to respond to, followed by a feedback stimulus indicating whether performance was accurate. During each trial, the cue “X” was shown for 500ms, followed by a fixation point for 11,500ms. Depending on subjects’ performance the feedback stimulus was then shown for 500ms, followed by yet another fixation point for 11,500ms. The feedback stimulus was a plus sign “+” to indicate performance was correct, and it was a minus sign “−” to indicate a performance error.

Participants were instructed to make a 1-s duration estimate by making a right index finger button press following onset of the cue as soon as they thought that 1s had passed. As in the report by Miltnner et al. (1997), responses were judged to be correct when the button press fell within a varying time window centered around 1000ms following cue onset. This time window can be described as the interval \([1000\text{ms}−t; 1000\text{ms}+t]\), whereby \( t \) varied as a function of performance. When responses were too fast or too slow, \( t \) was increased with 20ms making correct feedback more likely; when responses were correct, \( t \) was decreased with 20ms making error feedback more likely. The initial value of \( t \) was 150ms. As a result, the window length adjusted itself such that each participant received approximately as many correct as error feedback stimuli independent of their performance.

Next, participants performed three blocks of a button-press version of the Stroop task. Before entering the magnet, subjects were trained on the response mapping. During this training they performed one block of trials, having been instructed to read the word and respond using a manual response. This would increase the mapping of the word to the response finger, thus increasing the amount of errors and response conflict (e.g., Ruff, Woodward, Lauren, & Liddle, 2001). Participants were instructed to make a left index finger button press when the word was RED or YELLOW, and a right index finger button press when the word was GREEN or BLUE. They were instructed that inside
the scanner, they would have to respond to the color rather than the word, and were reminded again prior to the onset of the Stroop task.

During congruent trials, the color and the word were mapped onto the same response finger (e.g., RED in the color red or YELLOW in the color red, both requiring a left-hand response). During incongruent trials, color and word were mapped onto opposite response hands. Participants performed three blocks of 124 trials each, each block started with a fixation point lasting for 12s. The first four trials in each block were congruent, for the rest stimuli were presented in random order with 75% being congruent and 25% incongruent. Each stimulus lasted 300ms, and was followed by a 2700ms fixation point. Following the last trial in each block, another 12s fixation was presented. Participants were instructed to respond fast but accurately. Stimuli were presented using E-Prime (Psychological Software Tools, Pittsburgh, PA).

2.3. MR image acquisition and analysis

MR images were acquired using a 3.0 Tesla scanner (General Electric Company) with a standard head coil, using a spiral pulse sequence (TR = 1.5s; TE = 181ms; FOV = 20cm; flip angle = 70°). We acquired 28 functional images parallel to the AC-PC line, with the 23rd slice from the top through the middle of the AC. Data analysis was performed using BrainVoyager software (Brain Innovation, Maastricht, the Netherlands). Functional data was preprocessed using 3D motion correction, interscan slice time correction, Gaussian spatial filtering (FWHM = 6mm), and high-pass filtering (low cutoff: 3 cycles/block, or .0078125Hz). In addition, for the time estimation task, data were smoothed using a 2.8 s FWHM Gaussian filter; this preprocessing step was omitted in the analysis of the data from the Stroop task, because this was a fast event-related design, and temporal smoothing is thought to be suboptimal for such designs (Della-Maggiore, Chau, Peres-Neto, & McIntosh, 2002). For each subject, 3-dimensional anatomical images (SPGRs) were acquired; functional data were aligned to these and then transformed into Talairach space. For the time estimation task, statistical analysis was performed using a general linear approach using subject as a random factor with three predictors: One predictor accounted for the cue, one for correct feedback, and one for error feedback. For the Stroop task, statistical analysis was performed using a general linear approach, also using subject as random factor, with three predictors: one predictor accounted for congruent trials (correct trials only), one for incongruent trials (correct trials only), and one for errors. These predictors were obtained using a model of the hemodynamic response to the onset of each trial type. A statistical threshold of \( p < .005 \) was used, along with a contiguity threshold of 100 mm³.

3. Results

3.1. Time estimation task performance data

RTs were about as fast on correct judgments (\( M = 986 \) ms, \( SD = 43 \)) as on incorrect judgments (\( M = 974 \) ms, \( SD = 209 \)), Wilcoxon signed ranks \( z(13) = .85, \) ns. Also, participants went through about as many correct trials (\( M = 33 \) trials, \( SD = 3.25 \)) as error trials (\( M = 31 \) trials, \( SD = 3.25 \)), \( t(13) = 1.64, ns \), replicating the findings by Miltner et al. (1997). The average window width on the basis of which responses were judged to be correct or not was ±132 ms. This is also similar to the study by Miltner et al.; the average window widths in their study were ±132 ms in the auditory condition, ±138 ms in the somatosensory condition, and ±124 ms in the visual condition.

We analyzed whether subjects used the feedback to guide their time estimations in the same way that Miltner et al. (1997) analyzed their data. That is, we calculated the absolute RT difference between each trial \( n \) and subsequent trial \( n + 1 \) (ignoring, of course, the last trial of each block). Then we analyzed that difference as a function of whether subjects had received correct or error feedback on trial \( n \). Replicating the findings by Miltner et al. (1997), participants changed their time estimates more following error feedback (\( M = 202 \) ms, \( SD = 86 \)) than following correct feedback (\( M = 126 \) ms, \( SD = 49 \)), Wilcoxon signed ranks \( z(13) = 3.23 (p = .001) \). These results strongly suggest that participants performed the task in a similar way as the participants of Miltner et al. did; they used the feedback to guide and adjust their behavior.

3.2. Stroop task performance data

Mean RTs in the congruent condition (\( M = 612 \) ms, \( SD = 151 \)) were significantly faster than RTs in the incongruent condition (\( M = 719 \) ms, \( SD = 203 \)), \( t(13) = 6.5, p < .001 \). Mean accuracy for the congruent condition (\( M = 94\% \), \( SD = 7 \)) was significantly higher than for the incongruent condition (\( M = 85\% \), \( SD = 12 \)), Wilcoxon signed ranks \( z(13) = 2.94, p = .003 \). These findings replicate typical Stroop results (MacLeod, 1991): incongruent trials elicit slower and less accurate responses.

3.3. fMRI data

First, we investigated whether error feedback stimuli in the time estimation task elicited a greater hemodynamic response than correct feedback stimuli. No area in the brain displayed this result (see Table 1), not even at a decreased statistical threshold (\( p < .2 \)). Results did
show (see Fig. 1A–C) that correct feedback stimuli, compared to error feedback stimuli, engaged a distributed network of brain areas consisting of the bilateral caudate nuclei, the right putamen, the right rostral and posterior cingulate cortices, bilateral middle and inferior prefrontal gyri, bilateral superior temporal gyri, and the right lateral and medial occipital cortex (see Table 1, Fig. 1).

To verify whether we were simply underpowered or otherwise unable to detect any activity of the ACC to error feedback, we investigated whether response conflict and action slips in the Stroop task led to ACC activity. We will only report ACC activity here; additional analysis of the results of this Stroop task will be reported elsewhere (van Veen et al., submitted). Two contrasts were analyzed: incongruent versus congruent trials, and errors versus correct trials. Results of the first contrast did include a region of the caudal ACC (485 mm$^3$; $x = 0$, $y = 25$, $z = 28$). Results of the second contrast resulted in three active regions in the ACC. One was located caudally (379 mm$^3$, $x = 1$, $y = 26$, $z = 33$), and overlapped with the area engaged by response conflict (see Fig. 1D). A second region was located more rostrally (347 mm$^3$, $x = 1$, $y = 40$, $z = 22$). A third region was located more posterior and dorsally, and extended into the supplementary motor areas (3672 mm$^3$, $x = 4$, $y = 7$, $z = 58$). These results clearly show that our participants were quite able to activate their ACC, and that we were quite able to detect this activation.

A last set of tests involved performing post-hoc statistics using paired $t$ tests on the three ACC regions of interest (ROI) taken from the Stroop test. For the rostral ACC region, beta estimates were 5.03 ($SD = 8.53$)
for correct feedback and 2.98 (SD = 5.12) for error feedback, \((t(13) = .7, ns)\). For the caudal ACC region overlapping with the conflict area, beta estimates were 7.88 (SD = 3.94) for correct feedback and 6.08 (SD = 5.45) for error feedback, \((t(13) = 1.44, ns)\). For the caudal ACC region extending into the SMA, beta estimates were 10.11 (SD = 5.63) for correct feedback and 8.01 (SD = 5.74) for error feedback, \((t(13) = 1.26, ns)\). Thus, for each ROI, activity was in fact higher to correct feedback than to error feedback, although not significantly so. This, again, suggests we were not merely underpowered; the fact that the signal was in the opposite direction argues against this possibility.

3.4. Discussion

Performance data clearly replicated the findings obtained by Miltner et al. (1997); the adjustable time window ensured that subjects received approximately equal amounts of correct and error feedback. Moreover, following error feedback, participants adjusted their behavior. These results strongly suggest that participants performed the task in a similar way as the participants of Miltner et al. did. Nevertheless, we did not observe increased activation of the ACC to error feedback in the absence of conflict that might be caused by expectancy violation, even at extremely low statistical thresholds. This contrasts with previous results using dipole modeling of the error feedback negativity (Miltner et al., 1997). Because our task design was almost identical to that used by Miltner et al., as were our performance results, we expect similar neural processes to be engaged during our task and during the study of Miltner et al. However, we failed to provide support for the dipole models obtained in that study. This result suggests that the error feedback negativity may be generated by areas other than the ACC.

It is, of course, possible that we were simply underpowered to detect error feedback-induced ACC activity. Typically, the ERN is a very large signal, and the error feedback negativity is much smaller, so it is possible that the ACC signal following error feedback was simply too small to be reliably detected using fMRI. However, there are several reasons why we believe that this is not likely to be the case. During the Stroop task, which subjects performed during the same scanning session, both response conflict and errors engaged the ACC (see Fig. 1D), replicating earlier findings showing that response conflict and errors engage the same area in the ACC (Carter et al., 1998; Kerns et al., 2004). It is notable that in ERPs, ACC activity during correct incongruent trials in the Stroop task is thought to be reflected in the so-called N450 (Liotti, Woldorff, Perez, & Mayberg, 2000; West, 2003; West & Alain, 2000), and this component appears to have a similar amplitude as the error feedback negativity does. Thus, if the error feedback negativity reflects ACC activity, this activity should be similar in magnitude to ACC activity during correct incongruent Stroop trials. Since we were not underpowered to detect ACC activity during such correct incongruent Stroop trials, we believe that it is unlikely we were underpowered to detect such activity associated with negative feedback.

Analysis of data obtained from the Stroop task indicated that slips activated three areas in the ACC: one rostral, one caudal and overlapping with an area engaged by response conflict, and one area caudal and extending into the supplementary motor areas. Post-hoc tests were performed on these three areas to investigate whether one of them would respond to error feedback. However, results from these tests indicated that, though not significant, the ACC activation was in fact larger following correct feedback than following error feedback. The fact that activation was not even in the right direction is another argument against the possibility that we were simply underpowered.

Results did, however, show a relative increase in activity of the basal ganglia to correct feedback (see Figs. 1A and C), consistent with research showing basal ganglia activity to rewarding feedback stimuli (e.g., Delgado, Nystrom, Fissell, Noll, & Fiez, 2000; Knutson, Westdorp, Kaiser, & Hommer, 2000). This is consistent with the RL theory of Holroyd and Coles (2002), which predicts increased basal ganglia activity following unpredicted positive or rewarding events.

As noted in the introduction, in several previous fMRI reports of ACC activity elicited by error feedback (Holroyd et al., 2004; Monchi et al., 2001; Ullsperger & von Cramon, 2003), a broad conflict-based explanation of this activity in these studies cannot be ruled out. For instance, Monchi et al. (2001) observed ACC activity to error feedback in the WCST, which clearly involves task set reconfiguration or expectancy violation. Ullsperger
and von Cramon (2003) and Holroyd et al. (2004) used rather different paradigms than the present one; in their tasks, participants had to make a two-button forced choice response, and feedback was provided shortly following the response. It is therefore possible that the feedback stimulus evoked the tendency to make a corrective response, leading to the ACC activation. If this is the case, this ACC activation with feedback might either reflect either the premotor functions of this region of the brain (low levels of activation are seen during compatible trials in response interference tasks), or conflict between engagement of the corrective response and an online representation of the executed response maintained in working memory (cf. Jeannerod & Frak, 1999). One way to test these possibilities would be to replicate the tasks used by Ullsperger and von Cramon, or Holroyd et al., measuring muscle force or the lateralized readiness potentials while varying the response-feedback SOA. It is unlikely that something similar to this would have occurred in the present study and in the study by Miltner et al. (1997); first, the response-feedback SOA was too long in our study, and second, the accuracy of the response was determined by its timing, so that it would not be possible to “correct” it. Finally, Bush et al. (2002) observed increased ACC activity following feedback indicating reduced reward. They had participants press one of two buttons following a cue, after which they received a feedback stimulus. In 80% of trials, participants received a “reward” feedback stimulus, indicating they earned 15 cents, and had to push the same button on the next trial. In 10% of trials, the feedback stimulus instructed the participants to switch, that is, push the other button on the next trial. In the remaining 10% of trials, they received a “reduced reward” feedback, indicating that they earned only 9 cents that trials, and had to push the other button on the next trial. ACC activity was highest to the reduced reward feedback, somewhat lower to the switch feedback, and lowest to the constant reward feedback. Based on these results, Bush et al. proposed that the ACC plays a role in what they refer to as “reward-based decision making,” rather than merely responding to the amount of conflict that occurs in the information processing stream. Note, however, that conflict theory can still account for these findings. Trials during which the participants received a stimulus instructing them to switch occurred on 20% of the trials (switch trials plus reduced reward trials), while participants’ reward was reduced on 10% of the total amount of trials. Thus, the increased ACC activity during these trial types might be considered as related to novelty; the same goes for other studies (Monchi et al., 2001). Alternatively, it is possible that the ACC activity during such trials still reflected response conflict, between the response established under previously unambiguously rewarded conditions and the preparation of the alternative response. Because of these ambiguities in the literature we believe that it remains an open question as to how feedback engages the ACC and whether it does so independently of the conflict which it may elicit. Further studies will be needed to resolve this question.

It is worth emphasizing again that the concept of conflict, by the present account, goes beyond response conflict. In the papers by Botvinick et al. (2001), Jones et al. (2003), and Yeung et al. (in press), conflict was defined as Hopfield’s energy calculated over the response layer of a PDP network. Previous studies have indicated that the ACC appears to respond selectively to response conflict (Milham et al., 2001; Van Veen et al., 2001). More recent evidence, however, has suggested that the ACC appears to respond also to conflict between other types of representations as well (Badre & Wagner, 2004; Milham et al., 2003) including novelty or expectancy violation (Clark et al., 2000; Dien et al., 2003; Downar et al., 2000; Kiehl et al., 2001). In fact, Botvinick et al. theoretically left it an open question as to which kinds of conflict engage the ACC (Botvinick et al., 2001, pp. 630, 646). Perhaps novel modeling studies are needed to evaluate the implications of this slightly expanded view, to generate novel predictions and to prevent the theory from becoming unfalsifiable.

As stated earlier, the fact that we failed to provide support for the dipole models by Miltner et al. (1997) suggests that the error feedback negativity is generated by areas other than the ACC. However, in our study, no brain area was engaged to a stronger degree by error feedback than by correct feedback. Because ERPs cannot distinguish between relative increases or decreases in activity, it is possible that the error feedback negativity in fact reflects the increased engagement of one or more of the areas to correct feedback. Perhaps, then, the error feedback negativity is better conceived of as a superimposed “correct feedback positivity.” One possible way to test this would be to investigate correlations between the amplitude of participants’ error feedback negativity and the fMRI signals associated with positive and error feedback. The hypothesis outlined above predicts that the amplitude of the correct-error feedback difference wave should be correlated with correct feedback-related activity of a generator somewhere in the brain other than the caudal ACC.

In summary, in the present study using event-related fMRI and a long delay we were unable to provide evidence that the ACC is involved in error feedback-detection. This finding therefore fails to provide support for the RL theory of ACC function and the dipole models on which it is, in part, based. It is therefore possible that the error feedback negativity is not generated by the ACC but by other elements of systems evaluating performance and feedback. But while this negative result
appears consistent with conflict theory, it is based upon a null finding. It is conceivable that the small differences between the current study and the study by Miltner et al. (1997) were responsible for the lack of results. A possible factor is the difference in timing; the current study used a very long cue to feedback SOA, and it is possible that, despite the fact that performance measures indicated that participants executed the task in the same way, they processed the feedback in a different way. Another possibility is that ACC activity might only come into play at shorter response-feedback SOAs, either because there is less salience for delayed feedback or because there is less conflict. If this is true, then we would expect the error feedback negativity to disappear with increasing response-feedback SOAs and the systematic manipulation of salience and conflict in such an experiment might help to elucidate the mechanisms underlying this response. A third possibility is that there is no one-to-one correspondence between neural activity recorded at the scalp (as measured in the Miltner et al. study) and blood flow (as measured in the present experiment), thus, it is possible that the ACC activity that produces the error feedback negativity cannot be picked up by fMRI. However, this seems somewhat unlikely, as we were able to measure ACC activity on errors and conflict trials during the Stroop task. Lastly, it is possible, even likely, that generation of the error feedback negativity is a more complex process, to which multiple brain areas might contribute; contribution of the ACC to this process might be more paradigm-dependent. Therefore we conclude that the involvement of the ACC in processing error feedback remains an open question that merits further careful investigation.

Acknowledgments

This work was supported by MH64190 from the NIMH and a Translational Clinical Scientist Award by the Burroughs-Wellcome Foundation to Dr. Carter. We thank Mylaina L. Gordon for helping with the data collection.

References

Gruber, O., & Goschke, T. (2004). Executive control emerging from dynamic interactions between brain systems mediating language,


