Psychological Review
2004, Vol. 111, No. 4, 931–959

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.4.931

The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity

Nick Yeung
Princeton University

Matthew M. Botvinick
University of Pennsylvania

Jonathan D. Cohen
Princeton University and University of Pittsburgh

According to a recent theory, anterior cingulate cortex is sensitive to response conflict, the coactivation of mutually incompatible responses. The present research develops this theory to provide a new account of the error-related negativity (ERN), a scalp potential observed following errors. Connectionist simulations of response conflict in an attentional task demonstrated that the ERN—its timing and sensitivity to task parameters—can be explained in terms of the conflict theory. A new experiment confirmed predictions of this theory regarding the ERN and a second scalp potential, the N2, that is proposed to reflect conflict monitoring on correct response trials. Further analysis of the simulation data indicated that errors can be detected reliably on the basis of post-error conflict. It is concluded that the ERN can be explained in terms of response conflict and that monitoring for conflict may provide a simple mechanism for detecting errors.

Errors are an important source of information in the regulation of cognitive processes. The mechanism by which people detect and correct their errors has been the object of study for many years, but research interest has increased in recent years following the discovery of neural correlates of performance monitoring. In particular, studies of event-related brain potentials (ERPs) have revealed a neural response following errors that has been labeled the error-related negativity (*ERN* or *Ne*; Falkenstein, Hohnsbein, Hoorman, & Blanke, 1990, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). The most likely neural generator of the ERN is anterior cingulate cortex (ACC), an area that in recent years has been implicated in another function related to the evaluation of performance, monitoring for competition (or *conflict*) during information processing. The present research attempts to provide an integrative account of error- and conflict-related activity observed in anterior cingulate cortex. Specifically, we propose a new account of the ERN and error processing in terms of the conflict monitoring theory of anterior cingulate function.

Nick Yeung, Department of Psychology, Princeton University; Matthew M. Botvinick, Department of Psychiatry and Center for Cognitive Neuroscience, University of Pennsylvania; Jonathan D. Cohen, Department of Psychology, Princeton University, and Department of Psychiatry, University of Pittsburgh.

Preliminary versions of this research were presented in poster format at the conference on Executive Control, Errors, and the Brain (Jena, Germany, September 2000), at the 7th International Conference on Functional Mapping of the Human Brain (Brighton, England, June 2001), and at the 31st Annual Meeting of the Society of Neuroscience (San Diego, California, November 2001). We thank Mike Coles, Clay Holroyd, John Kounios, Sander Nieuwenhuis, and Leigh Nystrom for comments on drafts of this article; Joe Bussiere, Jack Gelfand, Mike Scanlon, and Dan Zook for help in running the event-related brain potential experiment; and Richard Greenblatt for technical advice.

Correspondence concerning this article should be addressed to Nick Yeung, who is now at the Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: nyeung@cmu.edu

Background

Behavioral Studies of Error Monitoring

Participants in reaction time (RT) experiments are typically aware of their errors, reacting to them with visible or audible frustration. When asked, they are also able to signal their errors more systematically with an appropriate key-press (Rabbitt, 1966, 1967, 1968). Using this method, Rabbitt and colleagues have found that participants can detect most, though rarely all, of the errors they make in simple choice RT tasks (Rabbitt, 1968, 2002). However, these error-signaling responses can be quite slow and unreliable. In a study by Rabbitt (2002), for example, young adults detected 79% of their errors, taking an average of about 700 ms to do so, when they were given a second to respond before the next stimulus appeared. However, when the subsequent stimulus appeared 150 ms after an incorrect response, the same participants showed a limited ability to ignore this stimulus, as they were instructed to do, and their error-detection rate dropped to 56%.

Whereas explicit error detection and signaling appear to be slow and effortful, error *correction* is fast and relatively automatic. Participants deal with errors more quickly and efficiently by producing a correcting response—that is, making the response they should have made—than by making a common detection response to all errors (Rabbitt, 1968, 1990, 2002). Indeed, errors are often immediately followed by a correcting response even when participants are instructed to suppress such responses (Maylor & Rab-

bitt, 1987; Rabbitt & Rodgers, 1977). These error-correcting responses can be extremely fast: Corrections are often observed within 20 ms of the original incorrect response (Rabbitt, Cumming, & Vyas, 1978), and in Rabbitt's (2002) study, the mean time to correct errors was around 250 ms.

Rabbitt and colleagues (Rabbitt et al., 1978; Rabbitt & Vyas, 1981) have explained fast, automatic error corrections in terms of the continuous flow of perceptual information into the response selection process (cf. C. W. Eriksen, Coles, Morris, & O'Hara, 1985; B. A. Eriksen & Eriksen, 1974; Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988). They describe response selection as involving accumulation of information over time, and they liken this to the votes of a committee. On occasion, they suggest, an incorrect decision will be made on the basis of incomplete information, but "as subsequent votes come in, a more accurate consensus will accumulate and the earlier mistake will become apparent" (Rabbitt & Vyas, 1981, p. 225). An implication of this hypothesis is that error correction (and detection) should depend crucially on continued information processing after the initial error. To test this hypothesis, Rabbitt and Vyas (1981) measured error correction rate as a function of stimulus duration: As stimulus duration is increased, and hence also the opportunity for further processing, the rate of error correction should increase, and this is exactly what Rabbitt and Vyas observed. Thus, participants' ability to detect and correct errors appears to be critically linked to their ability to continue processing the stimulus even after they initiate a response.

The Error-Related Negativity

The behavioral findings reviewed above have been complemented in recent years by data from neuroimaging techniques. In particular, a great deal of interest has focused on the ERN, a component of the ERP that is observed in association with incorrect responses (Falkenstein et al., 1990, 1991; Gehring et al., 1993). The term *error-related negativity* has in fact been used to label ERP components observed in at least three situations: following overt response errors in choice RT tasks (Falkenstein et al., 1990, 1991; Gehring et al., 1993); following feedback about response accuracy (Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997); and following late responses in deadline RT tasks (Johnson, Otten, Boeck, & Coles, 1997; Luu, Flaisch, & Tucker, 2000; Pailing, Segalowitz, & Davies, 2000). In this article, we are concerned with the issue of how the cognitive system is able to monitor its own performance in the absence of explicit feedback (e.g., regarding accuracy or timing). That is, we are concerned with the first of these situations, the observation of an ERP negativity immediately following incorrect responses. Henceforth we use the term *ERN* to refer specifically to this component. Thus defined, the ERN is a negative deflection in the ERP that begins around the time of incorrect responses, often slightly before, and peaks roughly 100 ms thereafter (see Coles, Scheffers, & Holroyd, 1998; Falkenstein, Hoorman, Christ, & Hohnsbein, 2000, for recent reviews). We return in the General Discussion to the relationship between this component and the related negativities observed following feedback and late responses.

Although the ERN varies in amplitude across experimental conditions, its latency seems to be very consistent (Falkenstein et al., 2000; Leuthold & Sommer, 1999). The ERN has been observed following errors regardless of the modality in which the stimulus is presented (Falkenstein et al., 2000) and regardless of the modality in which the response is made (Holroyd, Dien, & Coles, 1998). A number of features of the ERN have been taken to suggest that it indexes some form of error processing: Its amplitude is correlated with subjective judgments of response accuracy (Scheffers & Coles, 2000), is increased when response accuracy is emphasized over speed (Falkenstein et al., 1990; Gehring et al., 1993), and is reduced following incorrect responses to stimuli that are presented relatively infrequently, conditions in which errors are particularly likely (Holroyd & Coles, 2002). The ERN also appears related to aspects of error correction: Gehring et al. (1993) found that ERN amplitude correlates positively with the probability that an error is immediately corrected, and Rodríguez-Fornells, Kurzbuch, and Münte (2002) reported a larger ERN when an error is corrected quickly than when the error correction response is slow. There have also been attempts to correlate ERN amplitude with the force with which the error is produced: Gehring et al. (1993) found a negative correlation between these measures, with a larger ERN following weakly produced errors. However, a subsequent study found the opposite pattern, with larger ERN amplitude associated with more forceful errors (Scheffers, Coles, Bernstein, Gehring, & Donchin, 1996), a discrepancy that is not currently well understood.

A commonly held view is that the ERN reflects a monitoring process that signals errors whenever it detects a mismatch between the response produced and the correct, or *intended*, response, as determined by the state of the response system after the response is executed (Coles, Scheffers, & Holroyd, 2001; Falkenstein et al., 1990, 1991, 2000; Falkenstein, Hohnsbein, & Hoorman 1995; Gehring et al., 1993; Scheffers & Coles, 2000; Scheffers et al., 1996). This view is consistent with the error monitoring model proposed by Rabbitt and colleagues (Rabbitt & Rodgers, 1977; Rabbitt & Vyas, 1981). However, there is currently disagreement as to whether the ERN reflects the error-detection process itself (Coles et al., 1998; Falkenstein et al., 1991, 2000; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001; Scheffers et al., 1996), the arrival of the error signal at a remedial action system (Coles et al., 2001; Holroyd & Coles, 2002), or an emotional response to the error (Bush, Luu, & Posner, 2000; Gehring & Willoughby, 2002; Pailing, Segalowitz, Dywan, & Davies, 2002). Moreover, there is disagreement as to whether the representation of the correct or intended response depends on the final outcome of the response selection process (Falkenstein et al., 1990, 2000) or is determined by the state of the response system at the time of response execution (Coles et al., 2001).

The ERN has a frontocentral distribution that is symmetrical to the midline. Dipole modeling has consistently located its neural source in medial frontal cortex, consistent with a neural generator in ACC or the supplementary motor area (SMA; Dehaene, Posner, & Tucker, 1994; Gehring, Himle, & Nisenson, 2000; Holroyd et al., 1998). Convergent evidence favors ACC over SMA as the most likely source. First, recordings in behaving monkeys have found error-related activity in ACC and not SMA (Gemba, Sasaki, & Brooks, 1986; although Stuphorn, Taylor, & Schall, 2000, have observed error-related activity in the supplementary eye field during an eye-movement task). Second, fMRI studies have found error-related activity in ACC (e.g., Carter et al., 1998; Kiehl, Liddle, & Hopfinger, 2000; Menon, Adleman, White, Glover, & Reiss, 2001). Finally, it has been noted that the orientation of

pyramidal cells in the anterior cingulate sulcus could generate a frontocentral negativity such as the ERN, whereas SMA cells on the medial wall are oriented tangentially to the scalp and hence would not be expected to produce a corresponding scalp potential (Coles et al., 1998; Holroyd & Coles, 2002). Thus, these localization studies have led to the hypothesis that ACC is involved in detecting or responding to errors.

Response Conflict Monitoring

Although ACC activity has been observed in association with errors, fMRI studies have found that caudal ACC regions activated on error trials are typically also activated on trials when the participant responded correctly (e.g., Carter et al., 1998; Kiehl et al., 2000; Menon et al., 2001). More specifically, on trials with correct responses, ACC activity has been observed in conditions in which multiple responses compete for the control of action—that is, when there is *response conflict*. In the flanker task (B. A. Eriksen & Eriksen, 1974), for example, participants are required to make a discriminative response to a target stimulus, such as responding to *H* with the left hand and *S* with the right hand. The target is flanked by distractor stimuli that are associated either with the same response as the target (congruent stimuli, e.g., *HHHHH*), or with the opposite, conflicting response (incongruent stimuli, e.g., *SSHSS*). RTs and error rates are typically higher in the incongruent condition, the result of conflict during response selection between the responses afforded by the target and distractors—that is, between the correct and incorrect responses (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Gratton et al., 1988).

Botvinick, Nystrom, Fissell, Carter, and Cohen (1999) used a version of the flanker task in which participants were required to respond to the orientation of an arrow stimulus flanked by arrows pointing in the same direction (congruent stimuli, e.g., $<<<<<$) or in the opposite direction (incongruent stimuli, e.g., $<<>><<$). They observed ACC activity even on trials with correct responses, and this activity was greater for high-conflict, incongruent trials than for low-conflict, congruent trials. Findings such as these have led to the development of the conflict monitoring theory of ACC function (Botvinick, Braver, Carter, Barch, & Cohen, 2001; Botvinick et al., 1999; Carter et al., 1998). According to this theory, ACC is responsible for detecting conflict during response selection and conveying this information to brain regions directly responsible for the control of cognitive processing (e.g., lateral prefrontal cortex, Cohen, Botvinick, & Carter, 2000). Dealing with conflict, or crosstalk, in information processing has been proposed to be a central function of cognitive control (Allport, 1980, 1987; Neumann, 1987; Norman & Shallice, 1986): The presence of response conflict indicates situations in which errors are likely and, hence, in which attention is required. Thus, conflict monitoring may provide crucial information in regulating cognitive processing (Cohen et al., 2000). In addition, conflict monitoring is computationally straightforward, simply requiring the detection of concurrently active incompatible responses.

The conflict monitoring theory provides a unified account of neuroimaging findings of ACC activation in a wide range of task conditions associated with increased task difficulty (cf. Paus, Koski, Caramanos, & Westbury, 1998). For example, ACC is activated when participants perform the Stroop task (e.g., Bench et al., 1993; MacDonald, Cohen, Stenger, & Carter, 2000; Pardo, Pardo, Janer, & Raichle, 1990), when participants are required to produce infrequent responses in the face of more habitual ones (e.g., Braver, Barch, Gray, Molfese, & Snyder, 2001; Bush et al., 2000; Carter et al., 1998; Garavan, Ross, Murphy, Roche, & Stein, 2002; Kiehl et al., 2000; Menon et al., 2001; Paus, Petrides, Evans, & Meyer, 1993; Rubia et al., 2001; Taylor, Kornblum, Minoshima, Oliver, & Koeppe, 1994), and when they are required to choose between many valid responses in word generation tasks (Barch, Braver, Sabb, & Noll, 2000; Crosson et al., 1999; Thompson-Schill et al., 1997).

Conflict Monitoring and the ERN

The research reviewed above provides converging evidence that ACC is involved in some way in the evaluation of performance. However, the relationship between the error-detection function suggested by ERP data and the conflict monitoring function supported by fMRI studies remains a matter of debate. With regard to this issue, Carter et al. (1998) and Botvinick et al. (2001) have suggested that the conflict monitoring theory may be extended to explain ERP data as well as fMRI findings. Carter et al. (1998) noted that errors are particularly likely in conditions of response conflict, and they offered this as an account of ACC activity on error trials measured in electrophysiological recordings. Botvinick et al. (2001) later refined this hypothesis. In their connectionist model of conflict monitoring in the flanker task, the dynamics of response activation and conflict were very different on correct and error trials: Response conflict was larger on error trials than on trials with correct responses, particularly in the period following the response. Taking this finding of increased conflict following errors in a model of human performance, together with fMRI evidence that ACC is activated by conflict, Botvinick et al. suggested that the ERN may be explained by the response conflict monitoring theory.

The conflict monitoring theory thus promises to provide a unified account of ERP and fMRI findings concerning the role of ACC in performance monitoring. However, a number of objections have already been raised to the proposal that the ERN can be explained in terms of response conflict. A first criticism runs as follows: If the ERN reflects response conflict, then we should see an analog of the ERN—that is, a negativity following the response—on correct trials with high conflict. For example, one might expect there to be a larger negativity following correct responses on incongruent than congruent trials in the flanker task, because there is greater conflict on incongruent trials. However, such post-response negativities are typically not observed, and hence it is concluded that the conflict account of the ERN must be wrong (Pailing et al., 2000; Scheffers & Coles, 2000; Ullsperger & von Cramon, 2001).

Findings reported by Scheffers and Coles (2000) seem similarly difficult for the conflict theory to explain. These authors had participants perform the flanker task, but varied stimulus discriminability such that there were an appreciable number of errors to congruent as well as incongruent stimuli. When they measured the amplitude of the ERN as a function of stimulus congruence, they found a larger ERN following errors on congruent trials than on incongruent trials. Again, this result seems problematic for the conflict theory: The obvious expectation is that there should be more conflict, and hence a larger ERN, on high-conflict incongruent trials.

One goal of the present research is to address these empirical objections to the response conflict account of the ERN. Our strategy is to use a detailed model of the dynamics of response conflict in the flanker task to investigate how the conflict theory might explain these apparently troubling findings. To look ahead briefly, our simulation results demonstrate that, despite initial appearances, each of the empirical observations described above is entirely consistent with the conflict monitoring theory. Extending this investigation, we show that monitoring for response conflict could in principle provide a simple method for detecting errors. This demonstration begins to address a further objection to the conflict theory that is more theoretical in nature. That is, in seeking to explain the ERN in terms of conflict monitoring rather than in terms of an explicit error detection function, the conflict theory appears to leave unanswered the question of how participants are able to detect their errors (e.g., Rabbitt, 1966, 1968) and of why the ERN correlates with many aspects of human error processing (Falkenstein et al., 2000; Gehring et al., 1993). An aim of the present research is to demonstrate that the conflict monitoring theory can in fact provide answers to these questions. To this end, we demonstrate that conflict monitoring may provide a computationally simple method for detecting response errors.

Research Overview

In the present research, we develop the response conflict account of the ERN outlined by Botvinick et al. (2001) and extend this work to provide a new theory of how errors are detected in the brain. The research is presented in three sections.

Section 1. We first take a connectionist model of conflict monitoring in the flanker task previously developed in our laboratory (Botvinick et al., 2001; Cohen & Servan-Schreiber, 1992) and apply it to a range of ERN data. The aim is to demonstrate in a formally explicit manner how the conflict monitoring theory can explain the empirical phenomena of interest. We present five simulations concerned with:

1. The dynamics of response conflict on correct and error trials (Pailing et al., 2000; Scheffers & Coles, 2000; Ullsperger & von Cramon, 2001).
2. The effect of stimulus congruence on the ERN (Scheffers & Coles, 2000).
3. The impact of speed–accuracy instruction on the ERN (Falkenstein et al., 1990; Gehring et al., 1993).
4. Stimulus frequency and the ERN (Holroyd & Coles, 2002).
5. The relationship between ERN amplitude and error force (Gehring et al., 1993; Scheffers et al., 1996).

The first two simulations outline how the conflict monitoring theory deals with the two apparently troubling empirical findings described above. The third and fourth simulations demonstrate that our theory can account for other findings that are typically interpreted in terms of the properties of the error-detection system. The final simulation illustrates the utility of the model in providing insights into possible causes of discrepant results (concerning the ERN and error force).

Section 2. We next report the results of a new ERP experiment designed to test predictions of the conflict monitoring theory that arise from our simulations. To foreshadow the simulation results, an insight provided by the modeling work is that the dynamics of response selection and response conflict may be very different on correct and error trials. Specifically, our model suggests that error trials are characterized by response conflict in the period following the response, whereas when conflict occurs on correct trials, it is seen almost exclusively prior to the response. An implication of this point is that previous researchers may have failed to find ERP correlates of conflict monitoring on correct trials because they looked in the wrong latency range: Conflict-related activity should be observed prior to the response on correct trials, not in the latency range of the ERN (cf. Pailing et al., 2000; Scheffers & Coles, 2000; Ullsperger & von Cramon, 2001). We argue that the N2 component of the ERP (e.g., Pritchard, Shappell, & Brandt, 1991) is the electrophysiological correlate of this pre-response conflict on trials with correct responses. Section 2 presents an ERP experiment that tests predictions about the timing and neural source of the N2 that follow from this hypothesis.

Section 3. In the final section, we introduce a new theory of how error detection is implemented in the brain, based on the conflict monitoring theory. Although we propose that the ERN reflects conflict monitoring rather than a process that directly evaluates response accuracy, we do not intend to imply that the ERN is unrelated to error processing. Instead, we argue that conflict monitoring may provide a sufficient basis for detecting errors: Given that the response conflict model replicates many properties of the ERN, and that the ERN demonstrates many properties expected of an error-detection system, it seemed plausible to us that monitoring for response conflict might represent a simple method for detecting errors. In Section 3 we demonstrate that a conflict-based mechanism of error detection can perform with a reliability comparable to that exhibited by human participants in previous empirical studies.

1. The Response Conflict Theory of the ERN

In this section, we use a connectionist model of conflict monitoring in the flanker task to demonstrate that our theory can explain a variety of observed properties of the ERN. The use of a formal model allows us to explore in a principled way the properties and predictions of our theory. The benefits of this approach are twofold. First, the formal model makes explicit the structure and assumptions of the theory, allowing a more rigorous check of its internal consistency. Second, having a working model allows one to demonstrate the implications of the theory that, because of the complexities of the system described, may not be obvious on the basis of one's verbal theory or intuition alone. This property is particularly important because it allows the model to generate novel, testable predictions, and to suggest new explanations of existing findings.

Given that our theory of the ERN is based on the dynamics of response selection and response conflict, it is critical to have a good model of these dynamics. Fortunately, ERN researchers have typically used the flanker task described above, and the dynamics of response selection are perhaps better understood in this task

than in any other, following the work of Eriksen and colleagues (Coles et al., 1985; B. A. Eriksen & Eriksen, 1974; C. W. Eriksen et al., 1985; Gehring, Gratton, Coles, & Donchin, 1992; Gratton, Coles, & Donchin, 1992; Gratton et al., 1988). The empirical data generated by this research have led to the development of a computational model of response selection in this task that has been successful in accounting for findings from behavioral experiments (Cohen, Servan-Schreiber, & McClelland, 1992; Servan-Schreiber, Bruno, Carter, & Cohen, 1998; Servan-Schreiber, Carter, Bruno, & Cohen, 1998), ERP studies (Spencer & Coles, 1999), and fMRI studies (Botvinick et al., 2001). This model forms the basis for the present simulations.

Model Details

Model structure. The model is illustrated in Figure 1 and is described in more detail in the Appendix. It simulates performance in a task requiring a key-press response indicating whether the letter *H* or *S* appears in the center of a three-letter array, in which the flanking letters may be congruent or incongruent with the target letter. The basic model consists of three layers of units. There is an input layer consisting of an array of six position-specific letter units, a response layer with one unit for each response, and an attention layer with units corresponding to each location in the letter array. There are bidirectional excitatory weights between layers (information flow) and inhibitory links between all of the units within each layer (competition).

Botvinick et al. (2001) extended this model to implement the conflict monitoring theory of ACC function, adding a unit that is sensitive to the degree of conflict in the response layer. Conflict is calculated as the *energy* (Hopfield, 1982) of the response layer:

$$\text{Conflict} = -\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j w_{ij},$$

where a denotes the activity of a unit, w the weight of the connection between a pair of units, and the subscripts i and j are indexed over the units of interest. In the present model, the units of interest are the two response units, so the equation reduces to being the product of the activations of these units, scaled by the strength of the inhibitory connection between them. It is also important to note that each unit only sends activation to other units when its own activation level is positive, so that response conflict is effectively bounded at zero. Thus, when one response unit is active and the other inhibited, conflict is low or zero. However, when both units are active together, the product of their activations (and hence the degree of conflict) is large—capturing in a simple way the central notion of conflict. Thus, the level of conflict in the model is not a parameter of the model that is set by the experimenter. Rather, conflict is a measured property of the model; it depends only on the relative activation levels of the competing response units.

Although not crucial for the present purposes, Botvinick et al. (2001) used simulated ACC activity to vary the allocation of attention across trials. This feature of the model implements the proposal that ACC forms part of a control loop that is responsible for the flexible control of behavior and is not essential to the results reported here.

Model dynamics. Inputs corresponding to the four possible stimuli (the congruent stimuli, *HHH* and *SSS*; the incongruent stimuli, *HSH* and *SHS*) are simulated as patterns of activity across position-specific letter units of the model. When an input pattern is applied to the letter units, activation flows through their connec-

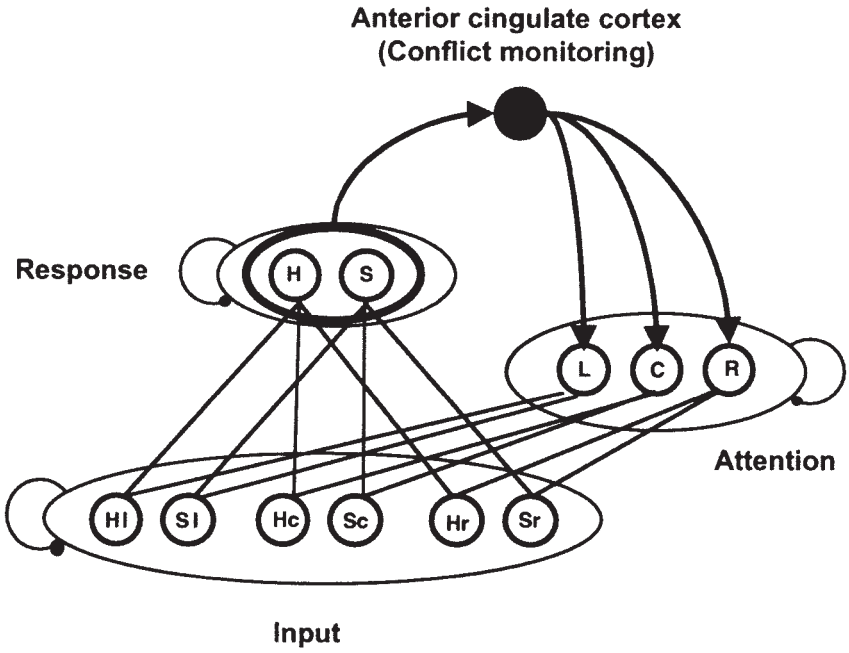

Figure 1. Illustration of the model of the flanker task used in the present simulations. The input, attention, and response layers are taken from the original model proposed by Cohen et al. (1992). The conflict monitoring feedback loop was added by Botvinick et al. (2001) to simulate the role of ACC in performance monitoring and adjustment of attentional control.

tions to the response units such that activation builds up in the response layer. A biasing input from the attention layer favors the letter in the center of the array, simulating the effect of visual attention to the target. Over the course of a few processing cycles, the network tends to settle into a state in which the target stimulus dominates the input layer and the corresponding response is activated in the response layer. The number of cycles required for the first response unit to reach a prespecified threshold is used to simulate RTs in the model. Following Botvinick et al. (2001), we calculate the simulated RT as:

$$\mathrm{RT(ms)} = 200 + (16 * \mathrm{cycles}).$$

The 200 ms constant is used to account for processes that are not part of the model. In particular, the model is intended to capture properties of the central response selection process, not the early perceptual processes that lead to stimulus identification. Spencer and Coles (1999) have shown that the model accounts well for ERP findings concerning response preparation in the flanker task, assuming that a response threshold crossing in the model corresponds to the onset of EMG activity. We follow this conclusion and, therefore, attribute the 200 ms constant to perceptual processes.

To simulate processing variability, we added noise to each unit at each time step (processing cycle). Because of this noise, the model occasionally responds before the stimulus is fully processed, simulating the impulsive responses that are observed empirically in this task (Cohen et al., 1992; Coles et al., 1985; Pailing et al., 2002). Impulsive responding leads to occasional errors. Such errors are particularly likely on incongruent trials because the flanking letters produce partial activation of the incorrect response unit that, together with noise in the system, push the activation of this unit above threshold. However, even on error trials, continued processing of the stimulus following the response, coupled with increasing attentional focus on the target letter, may eventually lead to activation of the correct response unit. If this activity is sufficient, the model will produce an error-correcting response. As becomes apparent, this tendency of the model to automatically correct its own errors provides the basis for our simulation of the ERN.

Simulation details. The results of the simulations reported here are based on 10 runs of 1,000 trials each, with randomized ordering of the stimuli. Except as noted, the parameters of the model are those used by Botvinick et al. (2001) to model behavioral and fMRI data. This is in line with our aim of investigating the qualitative features of the conflict theory of the ERN, rather than attempting to make detailed quantitative fits to specific data through an exhaustive parameter search. Where the model was run with different parameters to simulate the ERN in different experimental conditions, the parameters were chosen so that the performance of the model matched the relevant behavioral data. That is, parameters were not chosen to fit the electrophysiological data. Qualitatively similar patterns of results were found using a range of parameter values, demonstrating that the simulation results followed from the processing principles incorporated into the model rather than the particular parameters used.

Simulation 1: Response Conflict on Correct and Error Trials

The ERN is typically observed to begin around the time of error responses, often slightly before, and to peak around 100 ms later (see e.g., Coles et al., 1998; Falkenstein et al., 2000). The present simulation is concerned with this basic finding. To this end, Figure 2 shows the dynamics of response activation and conflict in the model on correct and error trials (averaging across congruent and incongruent trials). As is evident from the top panels of Figure 2, response conflict is larger and more sustained on error trials than on correct response trials. Figure 3 plots the difference in conflict between correct and error trials for the response-locked averages. The difference in conflict emerges around the time of the response and peaks 80 ms (five cycles) later. The post-error conflict signal thus replicates the primary features of the ERN. Henceforth, we refer to this difference in conflict between correct and error trials in the period immediately following the response as the *simulated ERN.*

Conflict on error trials. The simulated ERN can be explained in terms of the activation patterns of the two response units, which are given in the lower panels of Figure 2. As described above, the model produces errors when noise causes the incorrect response unit to cross threshold before the stimulus has been adequately processed (see Figure 2, lower panels). However, continued processing of the stimulus following the error causes the target stimulus unit eventually to dominate the competition between units in the input layer. As a consequence, activation of the correct response unit increases following the error (see Figure 2, middle panels). There is thus a brief period following incorrect responses

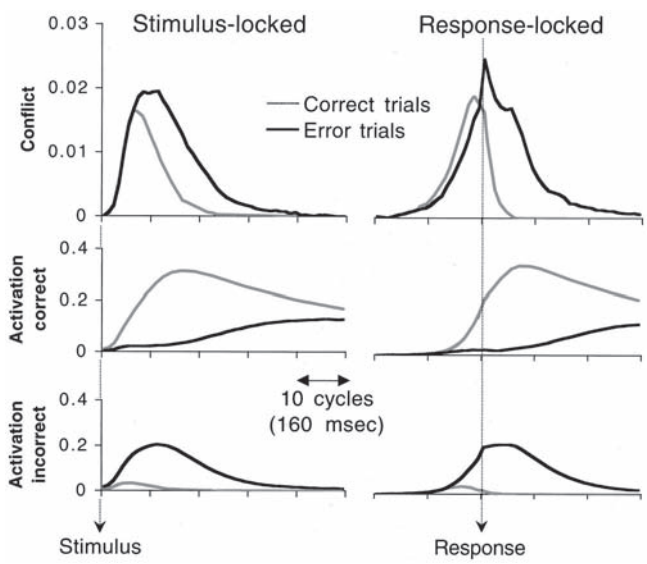

Figure 2. Activity in the network on correct and error trials of Simulation 1. Response conflict (simulated anterior cingulate cortex activity), upper graphs, is the scaled product of the activity in the correct response unit (middle graphs) and the incorrect response unit (lower graphs), bounded at zero. Left panels show the activity in the model averaged across trials aligned to stimulus onset. Right panels show corresponding response-synchronized averages, where trials are aligned with the response.

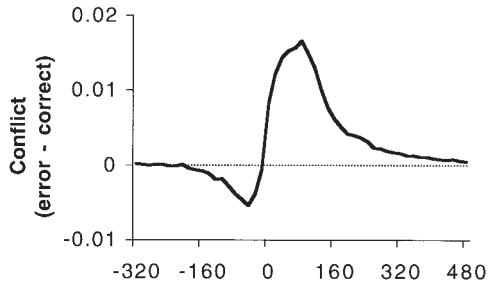

Figure 3. Difference in response conflict between correct and error trials of Simulation 1. The increased conflict on error trials in the period following the response (at time 0 ms) is the simulation of the error-related negativity.

during which both response units are activated, leading to a large conflict signal.[1]

A different pattern of activity is observed on correct trials. On these trials, activation of the correct response unit elicits the response and continued processing of the stimulus after the response simply reinforces the response decision made, with the correct response unit increasingly activated and the incorrect response unit becoming further inhibited (until external inputs to the network are removed). Response conflict is therefore largely restricted to the period prior to the response on correct trials, before inhibition from the correct response unit totally suppresses incorrect response activity. Thus, conflict following the response is observed only on error trials, and it is this post-error conflict that we associate with the ERN.

In summary, we explain the ERN in terms of response conflict that develops in the period following errors, a consequence of continued processing of the stimulus that leads to post-error activation of the correct response and hence conflict with the incorrect response just produced. A prediction of this hypothesis is that the ERN should be closely related to error-correcting activity. Gehring and Fencsik (1999) have reported empirical findings consistent with this prediction, showing that the ERN coincides with periods of coactivation of the correct and incorrect responses as measured through electromyography (EMG). More recently, Rodríguez-Fornells et al. (2002) have reported that the amplitude of the ERN is larger on trials for which the error-correcting response occurs quickly than on trials for which the error-correcting response is delayed, again consistent with the present theory.[2]

Conflict on correct trials. The simulation results suggest that conflict on correct trials is largely seen before the response. This point is illustrated further in Figure 4, which replots the simulation data separately for congruent and incongruent trials. For now, we draw attention to the pattern of response conflict observed on correct trials only (see Figure 4, upper middle panel, dotted lines). The difference in conflict between congruent and incongruent correct trials is largest in the period before the response, with little or no difference in conflict apparent afterward in the latency range of the ERN. Thus, the simulation results suggest that previous criticisms of the conflict theory may be misplaced: The theory does not predict that there should an analog of the ERN following the response on high-conflict correct trials (cf. Pailing et al., 2000; Scheffers & Coles, 2000; Ullsperger & von Cramon, 2001). Instead, the theory predicts that any negativity in the ERP associated with such conflict should be observed before the response. As discussed in detail below, we hypothesize that the N2 component of the ERP is the electrophysiological correlate of this pre-response conflict. This proposal forms the basis for the ERP experiment described in Section 2.

Because the simulation data show essentially no conflict in the period following correct responses, the model in its present form cannot account for recent observations of negative potentials following correct responses, with timing similar to that of the ERN (Vidal, Hasbroucq, Grapperon, & Bonnet, 2000). This may not be surprising: Coles et al. (2001) have argued that these correct-trial negativities may result from either evaluation of the timing of the response (cf. Luu et al., 2000) or from artifacts from stimulus-related negative components contaminating response-locked averages (cf. Vidal et al., 2000), neither of which is modeled in the present simulations.

[1] A salient feature of the response-locked data is the spike in conflict that occurs at the time of error commission (see Figure 2, top right panel). This spike is not observed in the stimulus-locked averages and can be understood in terms of a selection bias introduced by the interaction between noise variability in the simulations and the response-locking analysis procedure. Response-locking entails that activity in the incorrect response unit varies below threshold (by definition) at all points prior to error commission and varies above threshold (by definition) at the time of the incorrect response. Thus, activity in the incorrect response unit is always higher at the time of the response than on immediately preceding processing cycles, and hence, response conflict increases sharply at this point. Moreover, on processing cycles immediately following error commission, the high degree of activation of the incorrect response unit causes, via lateral inhibition, a temporary reduction in the activity of the correct response unit. As a consequence, conflict is reduced in the period immediately following the error.

[2] Coles et al. (2001) have reported findings relating the ERN to EMG activity that initially appear to present a challenge to our theory. They computed a measure of response conflict by multiplying together the maximum EMG activity of the two response hands, then selected subsets of correct and error trials that were matched according to this measure of conflict (Scheffers, 1999). Critically, even though the trials were matched for conflict in this way, ERN amplitude was greater on error trials than on correct trials, apparently challenging the present hypothesis. However, the method for calculating conflict used by Coles et al. may not be ideal. In particular, whereas response conflict is defined as the simultaneous activation of competing responses, Coles et al. estimated conflict by multiplying together two measures (maxima of EMG activity in the two response hands) that occur asynchronously. Moreover, in estimating conflict on the basis of peaks of EMG activity across entire trial epochs, their method does not specifically match for conflict in the post-response period that our theory associates with the ERN. Therefore, Coles et al.'s analysis method may not accurately match trials for the degree of conflict at the latency of the ERN. A reanalysis of the data from Simulation 1 produced results consistent with this interpretation. We followed Coles et al.'s procedure, selecting correct and error trials that were matched according to the product of activation maxima of the two response units. Although matched in this way, post-response conflict—the simulated ERN—was substantially larger following errors than following correct responses. These simulation results are consistent with the empirical findings and lead to a refinement of the prediction tested by Coles et al.: Our theory predicts that there should be no difference in ERN amplitude for correct and error trials that are matched for conflict in the post-response period when conflict is measured as the product of synchronous EMG activations.

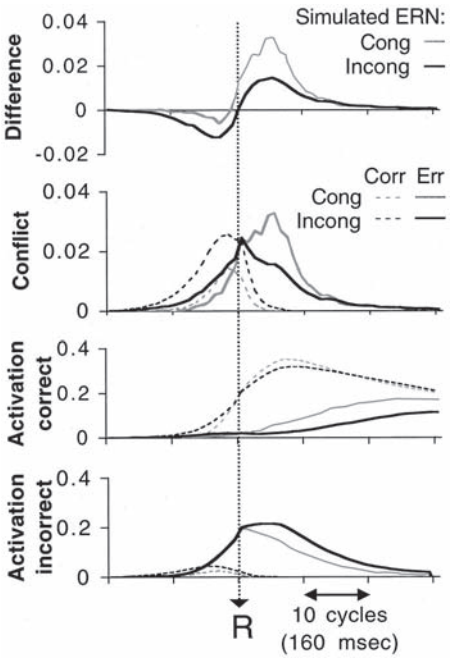

Figure 4. Results of Simulation 2, showing response-locked behavior of the model on congruent and incongruent trials, separately for correct and error responses. The top panel shows the simulated error-related negativity (ERN), calculated as the difference in conflict between the correct and error trials of each condition (shown in the panel below). The lower panels show activity in the correct and incorrect response units, from which the conflict measure is calculated. The vertical dotted line indicates the time of the response (labeled *R*). Cong = congruent; Incong = incongruent; Corr = correct trials; Err = error trials.

Simulation 2: ERN and Stimulus Congruence

The results of our second simulation demonstrate further the ability of the conflict monitoring theory to explain findings that initially appear troubling. The finding of interest is Scheffers and Coles's (2000) report of a larger ERN following errors to congruent stimuli than following errors to incongruent stimuli. At first glance, the conflict theory would appear to predict the opposite, because one might expect there to be greater post-error conflict, and hence a greater ERN, on high-conflict incongruent trials. However, a detailed simulation of the dynamics of response conflict on congruent and incongruent trials shows that the conflict theory in fact makes the opposite prediction, as shown in Figure 4. The top panel shows the simulated ERN—the difference in conflict between error trials and correct trials, shown in the upper middle panel—and the lower panels show activity in the correct and incorrect response units. An immediately striking feature of the simulation results is that post-error conflict is larger on congruent trials than on incongruent trials. That is, the simulated ERN is larger on congruent trials than on incongruent trials, with no difference in peak latency, consistent with the empirical findings of Scheffers and Coles (2000).

As described above, the simulated ERN is the result of post-error conflict between the incorrect response unit (that just led to the error response) and the correct response unit (that becomes activated because of continued stimulus processing). A crucial determinant of the amplitude of the simulated ERN is therefore the rate at which activation builds up in the correct response unit following the error. As is evident from the lower middle panel of Figure 4, there is more activation of the correct response unit on congruent error trials than on incongruent error trials—a straightforward consequence of the unambiguous nature of congruent stimuli. The result is a greater ERN on congruent trials. (It will be noticed also that activity in the incorrect response unit falls more quickly following the response on congruent trials than on incongruent trials. This lower level of post-response activity on congruent trials would tend to reduce the conflict signal. However, the reduction is proportionately smaller than the corresponding increase in post-response activity in the correct response unit.)

A further analysis of the simulation results demonstrates that the conflict theory can explain other aspects of Scheffers and Coles's (2000) results that initially seem troubling. In particular, Scheffers and Coles computed ERN amplitude for trials on which only one response was activated (as measured through EMG recordings). It seems that our theory might have difficulty in explaining why an ERN was observed at all on these trials, because only one response was activated. To address this issue, Figure 5 presents simulated ERN amplitude following congruent and incongruent stimuli for two sets of trials. The left-hand bars show the results for all trials (i.e., summarizing the results of Figure 4). The right-hand bars show the results for error trials on which correct response unit activation remained subthreshold throughout the trial. Evidently, for both congruent and incongruent conditions, a significant simulated ERN was observed even for trials in which error-correcting activity remained subthreshold. On these trials, the simulated ERN reflects conflict between the initial error and error-correcting activity that remains below the threshold for generating an overt response (as would be measured through EMG).

Simulation 3: ERN and Speed–Accuracy Instruction

Gehring et al. (1993) reported that the amplitude of the ERN is increased when accuracy is emphasized over speed, a result con-

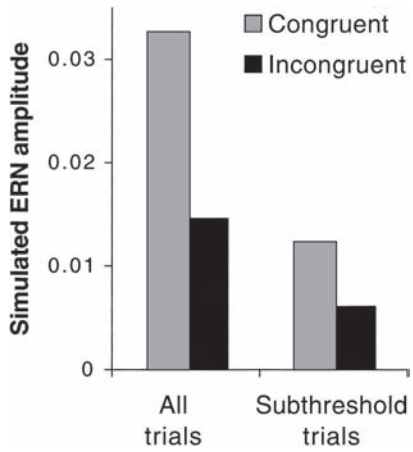

Figure 5. The amplitude of the simulated error-related negativity (ERN) on congruent and incongruent trials of Simulation 2. The results are shown for the average of all trials (left bars) and for the subset of trials in which activation of the correct response remained subthreshold throughout the trial (right bars).

firmed independently by Falkenstein and colleagues (Falkenstein et al., 1990, 2000). Gehring et al. accounted for their finding by proposing that errors are more salient to participants striving for accuracy than to participants for whom speed is the primary concern. That is, they explained their findings in terms of changes to the monitoring system—its sensitivity to the salience of errors—across conditions. It may not be initially obvious why a response conflict signal should be similarly sensitive to changes in accuracy instruction. However, a detailed simulation of Gehring et al.'s experiment not only demonstrates that the conflict monitoring theory can explain the empirical results, but also shows that our theory can provide a new explanation of these findings.

To model the results, we assumed that speed–accuracy instructions have two effects on the way participants perform the task. Specifically, participants striving for accuracy will adopt a more strict response criterion and will focus attention more strongly on the target letter to prevent errors on incongruent trials. Three speed–accuracy conditions were simulated. In the neutral condition, we used the parameters described in Simulation 1. An accuracy condition was modeled by increasing the threshold on the response units (from 0.18 to 0.20) and increasing the external input to the center attention unit (by a factor of 1.5). A speed condition was modeled by reducing the threshold on the response units (to 0.16) and reducing the external input to the center attention unit (by a factor of 0.75). With these parameters, error rates in the neutral condition were twice as large as in the accuracy condition, and in the speed condition they were three times as large, matching the behavioral results of Gehring et al.'s (1993) experiment. By this choice of parameters, we do not imply that it is not necessary to pay attention to the central target when performing under speeded task conditions. Our point is simply that attention to the central target is more necessary under conditions in which accuracy is stressed; under speeded conditions, a less focused attentional state can be beneficial because it allows participants to use flanker information and thereby reduce RTs on congruent trials.

The simulation results are shown in Figure 6. Consistent with the empirical findings of Gehring et al. (1993), the simulated ERN varied as a function of speed–accuracy condition, increasing in amplitude with more accurate performance. Despite changes in amplitude, the time at which the simulated ERN reached its peak did not vary across conditions, again consistent with empirical data. The difference in simulated ERN amplitude across conditions results from differences in post-error activation of the correct response unit (see Figure 6, third panel): Error-correcting activity is strongest in the accuracy condition, intermediate in the neutral condition, and weakest in the speed condition. These differences in post-error activity in the correct response channel are the direct result of the parameters that implement the changes in speed–accuracy condition: Greater attentional focus in the accuracy condition leads to more rapid post-error build-up of activity in the correct response unit and, hence, a larger simulated ERN.

In this way, the response conflict model simulates empirically observed properties of the ERN, while suggesting a different set of mechanisms by which these properties arise. Specifically, the simulation results suggest that processing changes required to increase accuracy may directly and necessarily lead to an increased ERN. The model does not require an additional assumption that the ERN is modulated by the salience of errors to the participant.

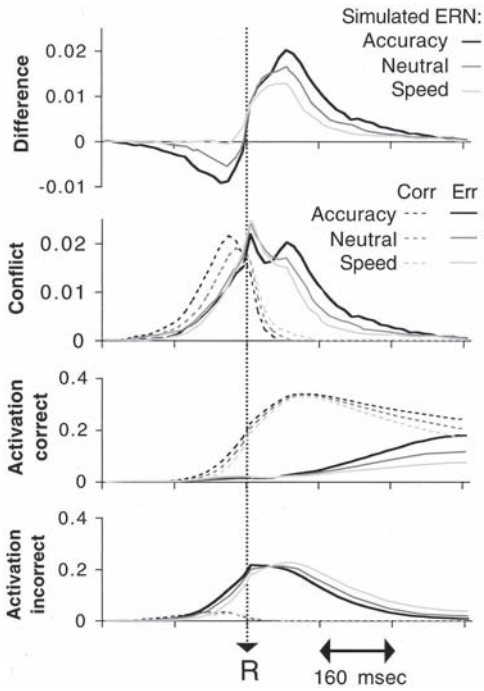

Figure 6. Results of Simulation 3, showing response-locked averages of network behavior separately for the accuracy, neutral, and speed instruction conditions. The top panel shows the simulated error-related negativity (ERN), calculated as the difference in conflict between the correct and error trials of each condition (shown in the panel below). The lower panels show activity in the correct and incorrect response units, from which the conflict measure is calculated. The vertical dotted line indicates the time of the response (labeled *R*). Corr = correct trials; Err = error trials.

Simulation 4: ERN and Stimulus Frequency

Holroyd and Coles (2002; Holroyd, Praamstra, Plat, & Coles, 2002) have studied the ERN in a modified version of the flanker task. Participants were presented with letter strings (*HHHHH*, *SSHSS*, *SSSSS*, and *HHSHH*), but stimuli with one target letter (e.g., *H*) were presented on 80% of trials, whereas stimuli with the other target letter (*S*) were presented on only 20% of trials. They labeled the highly probable target the *frequent* (F), and the less probable target the *infrequent* (I). The two letters were presented equally often as flankers. There were thus four conditions, infrequent congruent (III, 10% of trials), infrequent incongruent (FIF, 10% of trials), frequent incongruent (IFI, 40% of trials), and frequent congruent (FFF, 40% of trials). The participants were sensitive to the frequency manipulation, responding more quickly and accurately to frequent targets than to infrequent ones. There were too few errors to measure the ERN on frequent congruent trials (FFF), but otherwise ERN amplitude was found to vary across conditions: It was largest on frequent incongruent trials (IFI), intermediate on infrequent congruent trials (III), and smallest on infrequent incongruent trials (FIF).

Holroyd and Coles (2002) explained their findings in terms of their theory that the ERN is a signal indicating that the consequences of an action are worse than expected. In the case of infrequent incongruent trials, for example, participants made errors on nearly 70% of the trials. Holroyd and Coles suggest that

participants therefore expected to make errors in this condition, and hence, a small ERN was observed when they actually did so. In contrast, participants made errors on less than 5% of trials in the frequent incongruent condition (IFI). They therefore expected to be correct on these trials. When they were not, and hence the outcome was worse than expected, a large ERN was observed.

Holroyd and Coles (2002) therefore proposed that the error-monitoring system is sensitive to whether or not errors are predicted. The present model does not include any mechanism for predicting errors, yet a detailed simulation of response conflict in Holroyd and Coles's experiment demonstrates that our theory can explain their findings. To simulate the experimental results, we assumed that participants are sensitive to the relative probabilities of both stimuli and responses. In the experiment, one stimulus letter appeared in the target location four times as frequently as the other letter, and hence, one response was made at least four times as often as the other. We simulated the effects on processing of these differences in stimulus and response probabilities by increasing the gain of the stimulus unit coding the frequent target stimulus (by a factor of 1.5) and increasing the gain of the response unit coding the frequent response (by a factor of 1.2), with parameter values chosen so that the error rates in the model matched those found empirically by Holroyd and Coles. Thus, the model had a bias toward coding the target stimulus as the more frequent letter and a bias toward producing the more frequent response.

The results of the simulation are shown in Figure 7. The simulated ERN (see Figure 7 top panel) captures the pattern of the empirical data reported by Holroyd and Coles (2002): It is largest following errors on frequent incongruent (IFI) trials, intermediate on infrequent congruent trials (III), and lowest in the infrequent incongruent condition (FIF). As is evident from the lower middle panel of Figure 7, post-error activation of the correct response unit occurred most rapidly and most strongly in the frequent incongruent condition and was greatly reduced following errors to infrequent targets. This result is a straightforward consequence of the bias toward the frequent response used in the simulation. Thus, as in the previous simulations, the amplitude of the simulated ERN varied across conditions depending on how strongly the correct response unit competed with the activation of the incorrect response unit following the error.

Figure 8 plots simulated ERN amplitude as a function of response accuracy across conditions, contrasted with the empirical data obtained by Holroyd and Coles (2002; cf. their Figure 10). Although the model somewhat underestimates the ERN observed in the infrequent incongruent condition, the overall quantitative fit is good given that few parameter changes were made from the original model. As in the previous simulation, differences in simulated ERN amplitude across conditions are a direct consequence of changes in the dynamics of task processing according to task demands.

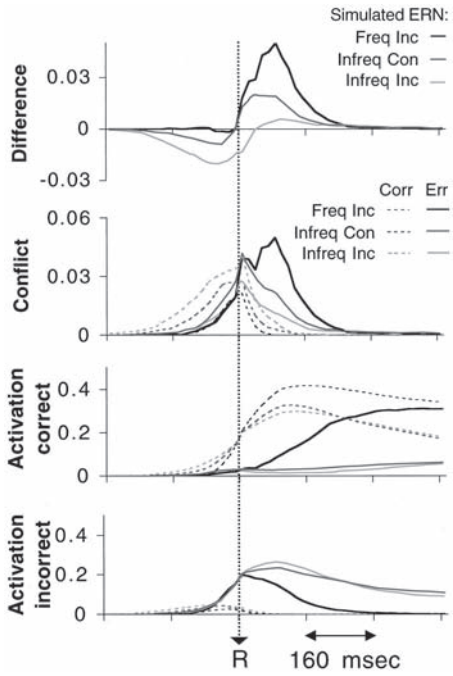

Figure 7. Response-synchronized averages for the frequent incongruent, infrequent congruent, and infrequent incongruent conditions of Simulation 4. The simulated error-related negativity (ERN), shown in the top panel, is derived from the conflict signals on correct and error trials that are given in the panel below. The lower panels show activity in the correct and incorrect response units. The vertical dotted line indicates the time of the response (labeled *R*). Freq Inc = frequent incongruent trials; Infreq Con = infrequent congruent trials; Infreq Inc = infrequent incongruent trials; Corr = correct trials; Err = error trials.

Simulation 5: ERN and Error Force

As mentioned briefly in the Introduction, attempts to relate ERN amplitude to the force with which errors are committed have produced seemingly contradictory results. Gehring et al. (1993) reported a negative correlation between these measures, with larger ERNs associated with smaller error force. By contrast, Scheffers et al. (1996) found the opposite pattern, with a larger ERN for motor (squeeze) errors than on trials with incorrect EMG activation but no overt error activity. There were many differences between the methods of Gehring et al. and Scheffers et al., including the experimental task used. However, of particular interest here are the differing rates of error correction observed in the two experiments. Gehring et al., using the flanker task, reported that roughly 70% of errors were corrected by their participants. Scheffers et al. used a hybrid choice–RT/go–nogo task. Participants responded to the direction of an arrow stimulus but were required to withhold responding on some trials depending on the orientation of a frame surrounding the arrow. Thus, participants could make "errors of action," responding on nogo trials—errors that could not be corrected with a further response. Participants could also respond with the wrong hand on go trials. It is these "errors of choice" that we are interested in here, and all such errors were corrected in their experiment.

The response conflict model predicts a very close relationship between the ERN and error correction. We therefore investigated whether the discrepant results of Gehring et al. (1993) and Scheffers et al. (1996) might be explained by the differing rates of error correction observed in their experiments. We defined error correction in the model as occurring when a threshold crossing in the incorrect response unit was followed by threshold crossing in the correct response unit. In Simulation 1, 63% of errors were corrected in this way, so the results of this simulation were taken as a reasonable model of the results of Gehring et al. To simulate the

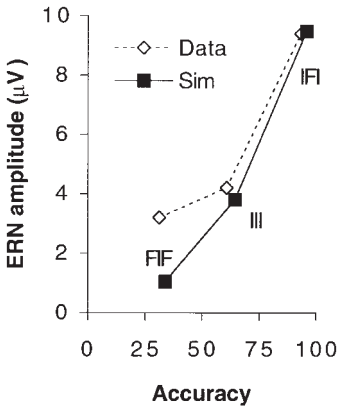

Figure 8. Error-related negativity (ERN) amplitude as a function of response accuracy across the stimulus conditions of Simulation 4. The results are shown alongside the empirical data from Holroyd and Coles (2002). The simulated ERN was fit to the data by multiplying the peak value of the simulated ERN (at Cycle 5 post-response) by 190. IFI = frequent incongruent trials; III = infrequent congruent trials; FIF = infrequent incongruent trials; data = empirical results; sim = simulation results.

results of Scheffers et al., we required parameters for which all errors would be corrected. This was achieved by increasing the external input to the center attention unit (by a factor of 1.25) and reducing the strength of the mutual inhibition between the two response units (from –3.0 to –1.5). This choice of parameters makes the general point that error-correction rates will vary as a function of at least two factors: the degree to which attention is effectively allocated to target information and the degree to which participants commit to their first response (captured by the level of inhibition between the two responses).

For each simulation, error trials were divided into four quartiles according to ERN amplitude. The error force on each trial was taken as the maximum activation in the incorrect response unit, and mean error force was calculated separately for each ERN quartile. The results are shown in Figure 9. The simulation with 63% of errors corrected replicates Gehring et al.'s (1993) finding of a negative correlation between ERN amplitude and error force, with ERN amplitude reduced on trials with strong activation of the incorrect response. The explanation for this finding is that trials with high error activity tend to be those on which the error is not subsequently corrected: The greater the activity in the incorrect response unit, the less likely it is that the correct response unit will overcome lateral inhibition and correct the error. Thus, trials with high error activation have little post-error activity in the correct response unit and hence little conflict. The result is a negative correlation between simulated ERN size and error force. By contrast, under simulation conditions in which all errors were corrected, a positive correlation between simulated ERN amplitude and error force is observed, replicating Scheffers et al.'s (1996) findings. In this case, post-error activation of the correct response unit occurs on every trial (because all errors are corrected). The primary determinant of ERN amplitude is therefore the degree of activation of the incorrect response unit: the greater the activation, the larger the conflict signal, and hence the positive correlation between error force and ERN amplitude.

The model is therefore able to explain apparently discrepant findings concerning ERN amplitude and error force and makes the general prediction that factors affecting the rate of error correction should influence the relationship between ERN amplitude and error force. Once again, our explanation of the empirical findings is in terms of the parameters in the processing mechanisms responsible for task performance, rather than the properties of a mechanism dedicated to error detection.

Discussion of Simulation Results

The conflict monitoring theory has previously been used to explain ACC activity observed in PET and fMRI studies. The aim of our simulations was to provide a formal investigation of the ability of this theory to explain observed properties of the ERN, a brain potential thought to be generated in ACC. According to our hypothesis, the ERN reflects conflict that develops in the period following errors. The simulation results demonstrate that this hypothesis can explain the timing of the ERN and its sensitivity to stimulus congruence, speed–accuracy instruction, stimulus frequency, and error force.

The simulation results first provided insight into the way in which our theory can explain findings that initially appear challenging. Simulation 1 showed that response conflict should be restricted to the period prior to the response on trials with correct responses, explaining why correlates of conflict monitoring are not observed after the response on correct trials (cf. Pailing et al., 2000; Scheffers & Coles, 2000; Ullsperger & von Cramon, 2001). Simulation 2 showed that conflict may be higher following errors on congruent trials than on incongruent trials because post-error activation of the correct response, and hence post-error conflict, is larger on congruent trials (cf. Scheffers & Coles, 2000).

The simulations also demonstrated that the conflict monitoring theory can offer alternative explanations of results that have previously been interpreted in terms of properties of the error-detection system. Thus, Gehring et al. (1993) explained their finding that ERN amplitude increases with response accuracy by proposing that "the ERN is associated with an error-related pro-

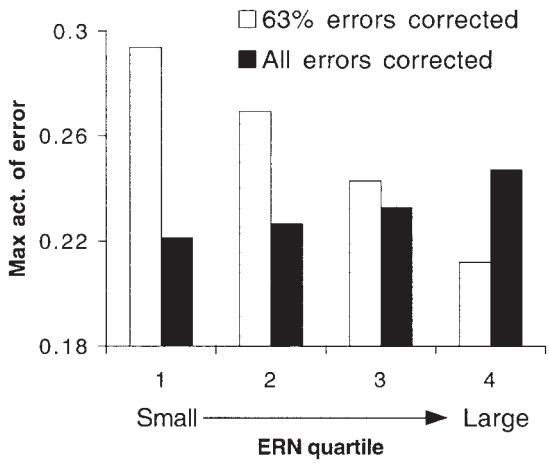

Figure 9. Results of Simulation 5, showing the error force associated with each error-related negativity (ERN) quartile as a function of the percentage of errors that were corrected. Max act. = maximum activation.

cessing system, whose activity is modulated by the degree to which accuracy is important to the subject" (p. 387). Meanwhile, Holroyd and Coles (2002) explained their finding of reduced ERN amplitude following errors to infrequent stimuli in terms of the sensitivity of the error-processing system to the expectedness of errors. By contrast, we explain both of these findings in terms of changes in task processing engendered by the task context. In Simulation 3, for example, we assumed that participants striving for accuracy would attempt to focus their attention more effectively on the central target letter. In the model, post-error activation of the correct response unit was increased as a consequence of increased attention, resulting in a high level of post-error conflict and a large simulated ERN. Of course, it may be possible to adapt existing theories to incorporate the idea that changes in task processing directly affect performance monitoring. Nonetheless, this possibility does not detract from the value of the present work in suggesting this novel account of the empirical data.

The final simulation demonstrated that the conflict monitoring theory can provide insight into the cause of apparently discrepant empirical findings. Gehring et al. (1993) found that ERN amplitude was negatively correlated with error force, leading them to suggest that the ERN reflects participants' attempts to "brake" the erroneous response—that is, they linked the ERN to mechanisms of error *compensation*. In contrast, Scheffers et al. (1996) found that ERN amplitude increased with error force, suggesting to them that the ERN relates to the magnitude of the error—that is, they linked the ERN to error *detection*. Our simulations offer a unified account of these findings, explaining the discrepant results in terms of the differing rate of error correction observed in the two experiments.

Overall, therefore, our model of conflict monitoring in the flanker task was able to simulate a range of empirically observed properties of the ERN, a formal demonstration that the conflict monitoring theory can explain these properties. The simulations were run without adjustments to the parameters associated with the monitoring process, reflecting the power of this simple mechanism to account for a wide range of empirical data. Furthermore, fits to empirical ERP data were achieved without reparameterization of the model for each simulation, evidence of the robustness of the results obtained. Nevertheless, an important goal for future research will be to refine the model in order to provide detailed quantitative comparisons with empirical data (and with competing theories when they are specified in comparable detail). However, perhaps the most stringent test of any model is its ability to generate testable predictions. In the next section we test one such prediction of the model, concerning ERP correlates of conflict monitoring on trials with correct responses.

2. ERP Correlates of Conflict on Correct and Error Trials

The simulation results suggest that response conflict should be limited to the period before the response on trials with correct responses. The simulation results therefore lead to the novel prediction that we should be able to measure conflict-related ERPs prior to the response on these trials. In fact, a good candidate for this conflict-related ERP, the N2, is already widely studied. In the flanker task, the N2 emerges around 250 ms after the presentation of the stimulus, has a frontocentral scalp topography, and is larger on incongruent trials than congruent trials (Heil, Osman, Wiegelmann, Rolke, & Henninghausen, 2000; Kopp, Rist, & Mattler, 1996; Liotti, Woldorff, Perez, & Mayberg, 2000). Kopp et al. (1996) have shown that N2 amplitude increases with the degree of activation of the incorrect response, as measured through EMG. Moreover, Liotti et al. (2000), using the Stroop task, localized the N2 component to ACC (see also Lange, Wijers, Mulder, & Mulder, 1998). The N2 therefore has all of the properties expected of a conflict-related ERP. However, none of the existing studies of the N2 observed in the flanker task have compared this component directly with the ERN, nor have they assessed the timing of the N2 with regard to the response. These questions are addressed in the present experiment.

The predictions tested in the experiment are as follows:

1. The N2 should be similar to the ERN in terms of scalp topography and neural source.

2. The N2 and ERN should differ in latency: The N2 should precede the response, whereas the ERN should follow it.

The first prediction is a straightforward implication of the theory that the ERN and N2 are both correlates of response conflict monitoring in ACC. The second prediction follows from our simulation results suggesting that conflict on correct trials should be limited to the period before the response, whereas error trials are characterized by post-response conflict.

Method

Participants. Nine female and 7 male undergraduate students from Princeton University (Princeton, NJ) participated in a single 2-hr session for course credit. All were right-handed and between 18 and 23 years old, and all had normal or corrected-to-normal vision. Informed consent was obtained from each participant at the start of the session.

Procedure. The participants were seated in front of a screen in a dimly lit room. They performed a version of the flanker task in which they responded by key-press to indicate the direction of a central arrow that was surrounded by flanker arrows. There were four stimuli, the congruent stimuli <<<<< and >>>>> and the incongruent stimuli << > << and >> < >>. The four stimuli were presented in pseudo-random order with the constraint that each stimulus appeared equally often in each block. On each trial, the participant was first presented with a fixation cross in the center of the screen. The cross was replaced after 500 ms with an imperative stimulus. The stimulus was presented for 100 ms and then the screen was cleared, remaining clear until 500 ms after the participant's response. At this time the string "- -" appeared to mark the intertrial interval, the duration of which varied randomly from 1,000 to 1,100 ms. All stimuli were presented in white on a black background. At a viewing distance of roughly 110 cm, the arrow stimuli each subtended 0.4° of visual angle vertically and 0.6° horizontally, and they were spaced 0.3° apart.

Participants first performed 2 or 3 practice blocks of 36 trials each. They then performed 12 blocks of 68 trials each during which behavioral and ERP data were collected. The participants were allowed to rest between blocks, at which time they were given feedback showing their mean correct RT and error rate in the previous block and for the whole session. If their error rate fell below 8%, they were encouraged to respond more quickly. If they made greater than 16% errors, they were told to respond more carefully. Participants were also encouraged through verbal instruction to sit in a relaxed position, to minimize eye movement, and to blink as seldom as possible while they performed the task.

Recording. The electroencephalogram (EEG) was recorded with Ag–AgCl electrodes from 64 locations arranged an extended 10–20 system montage in a fabric cap (Neurosoft, El Paso, TX), referenced to linked mastoids. The electrooculogram (EOG) was recorded from electrodes

placed above and below the left eye and to the sides of each eye to monitor eye movements. The ground electrode was placed on the chin. The electrode impedance for all electrodes was less than 50 kΩ. The EEG and EOG signals were amplified (Sensorium model EPA-6, Charlotte, VT; input impedance = 1 GΩ) by a gain of 20,000 with a 12-bit processor, filtered through a pass-band of 0.1–300 Hz (half-amplitude cutoff). The signals were digitized at 250 Hz.[3]

Data analysis. Stimulus- and response-synchronized epochs were extracted from the EEG off-line. Trials with blinks, large eye movements, instrument artifacts or amplifier saturation were rejected off-line through manual editing. For the ERN, we computed response-locked average waveforms for correct and error trials in an epoch beginning 200 ms prior to key-press and lasting 700 ms. The baseline window ran from –100 ms to 0 ms relative to the response. For the N2, we analyzed data only for trials with correct responses, computing stimulus-locked averages for congruent and incongruent trials separately. The epoch ran for 800 ms, beginning 200 ms prior to stimulus onset, with a 100 ms pre-stimulus baseline.

To compare the timing of conflict- and error-related ERP components, we extracted a second set of response-synchronized epochs from the EEG. These epochs ran from 800 ms prior to the response until 200 ms after, with the baseline period from 800 to 700 ms before the response. Response-locked waveforms were computed separately for correct congruent, correct incongruent, and error trials.

Component scalp topographies were analyzed using an analysis of variance (ANOVA) comparing voltage across 15 electrode sites (chosen to cover midline scalp areas known from previous studies to be the focus of the ERN and N2). Degrees of freedom were corrected using Greenhouse–Geisser epsilon values. Data from all 64 electrodes were then used in computing the most likely dipole source of each component. Dipole models were computed separately for the ERN difference wave (error – correct) and the N2 difference wave (incongruent – congruent).[4] Modeling was performed on unfiltered data, rereferenced to the average reference, across a 24-ms window around the component peak. The reported dipole solutions were stable across different seeding locations and were stable to the addition of further dipoles to the solution model. The validity of the dipole solutions was further assessed by applying them to the error grandaverage and incongruent correct grandaverage waveforms. These waveforms were first digitally high-pass filtered (>2 Hz) to remove the effects of slow parietal positivities seen around the time of the response. Fits to the resulting waveforms were comparable to those reported for the difference waveforms.

Results

Behavioral data. Mean correct RT was greater on incongruent trials than on congruent trials, averaging 421 ms and 352 ms, respectively, a reliable difference, $t(15) = 10.2$, $p < .01$. Error rates were also higher for incongruent stimuli (18.7%) than for congruent stimuli (2.1%), again a reliable difference, $t(15) = 10.5$, $p < .01$. Mean RT was higher on correct trials (386 ms) than on error trials (313 ms), $t(15) = 9.88$, $p < .01$. These findings are consistent with those of previous studies using the flanker task (e.g., Coles et al., 1985; B. A. Eriksen & Eriksen, 1974; C. W. Eriksen et al., 1985).

The ERN and N2. The upper panel of Figure 10 plots response-synchronized grandaverage waveforms for correct and error trials at electrode location FCz. An ERN is clearly evident as a negative deflection in the waveform on error trials that emerged just prior to the response and peaked 56 ms later. The ERN was followed by a large sustained positivity over posterior scalp regions, the error positivity (*Pe*; Falkenstein et al., 1995; Falkenstein et al., 2000). The lower panel of Figure 10 shows the scalp topography of the correct and error trial waveforms, along with the difference wave, at the time of peak ERN amplitude (56 ms after the response). A midline frontocentral topography is clear for error trials and for the difference wave.

To quantify the ERN, we performed a three-way repeated measures ANOVA using the average voltage in the 100 ms following the response. The factors were response accuracy (correct, error), anterior-posterior electrode location (F, FC, C, CP, P), and laterality (3, z, 4). A reliable main effect of response accuracy indicated that the waveform on error trials was more negative than that seen on correct trials, $F(1, 15) = 35.6$, $MSE = 99.3$, $p < .01$. That is, a robust ERN was observed. The ERN was largest at FCz and reduced in amplitude for electrodes away from this site, resulting in reliable interactions between response accuracy and anterior-posterior location, $F(4, 60) = 22.5$, $MSE = 4.0$, $\varepsilon = 0.45$, $p < .01$, and between accuracy and laterality, $F(2, 30) = 14.6$, $MSE = 4.0$, $\varepsilon = 0.94$, $p < .01$, and a reliable three-way interaction, $F(8, 120) = 4.4$, $MSE = 0.4$, $\varepsilon = 0.51$, $p < .01$. Taking the amplitude difference between correct and error trials as a measure of the ERN, pairwise comparisons revealed that the ERN was reliably larger at FCz than at the other electrode sites (all ps $< .01$).

The upper panel of Figure 11 plots stimulus-locked grandaverage waveforms for correct congruent and correct incongruent trials at electrode site FCz. An enhanced N2 was evident on incongruent trials. The difference between congruent and incongruent trial waveforms peaked 344 ms after stimulus onset. The lower panel of Figure 11 shows the scalp topography at this time for congruent trials, incongruent trials, and the difference between these conditions. A midline frontocentral topography is apparent for incongruent trials and for the difference wave. A smaller and more frontal negativity was present on congruent trials.

A three-way repeated measures ANOVA was performed using the average amplitude of the waveforms in a window running from 300 to 400 ms after stimulus onset. The factors were stimulus congruence (congruent, incongruent), anterior-posterior electrode location, and laterality. N2 amplitude was greater for incongruent trials than for congruent trials, $F(1, 15) = 25.19$, $MSE = 57.6$, $p < .01$. The amplitude of the congruence effect was largest at site FCz, indicated by reliable interactions between congruence and anterior-posterior location, $F(4, 60) = 5.3$, $MSE = 1.34$, $\varepsilon = 0.39$, $p < .05$, and between congruence and laterality, $F(2, 30) = 9.2$, $MSE = 0.7$, $\varepsilon = 0.96$, $p < .01$, although the three-way interaction did not reach significance, $F(8, 120) = 1.6$, $p > .1$. Using the amplitude

[3] The use of a high cutoff frequency (300 Hz) raises the possibility that aliasing of high-frequency signals contributed to our results. However, in system tests we found that signals above 30 Hz simply consisted of noise that was orders of magnitude weaker than the low-frequency signals of interest in the present research, suggesting that the effects of aliasing of high-frequency activity were negligible.

[4] Dipoles were fit using EMSE v4.2 (Source Signal Imaging, San Diego, CA) with a three-shell sphere model. Using the method described in Greenblatt and Robinson (1994), the algorithm used a three-shell sphere for each electrode to correct for nonspherical head shapes. The conductivity ratios for the sphere model were scalp:skull:brain = 1:0.0125:1 (Rush and Driscoll, 1968). The dipole fitting algorithm minimizes a chi-square cost function, using the iterative Nelder–Mead simplex algorithm (Press, Flannery, Teukolsky, & Vetterling, 1992) for the nonlinear components, with a minimum norm pseudoinverse solution for the linear components at each step.

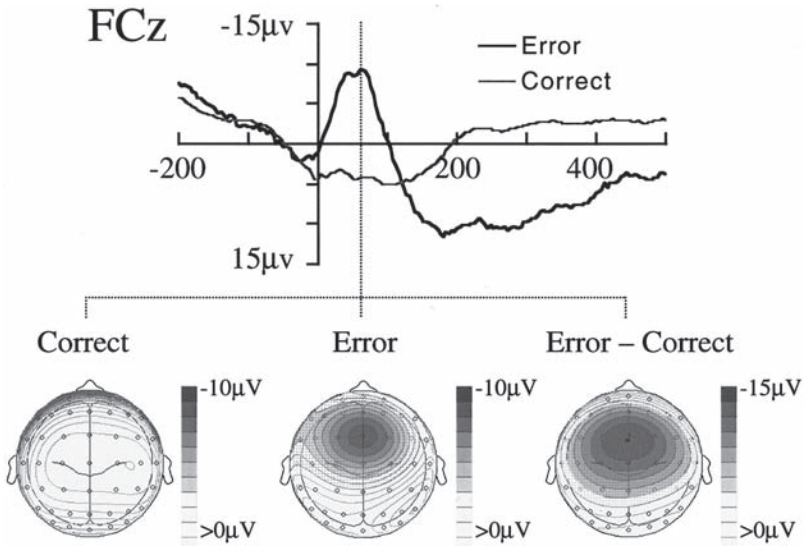

Figure 10. The error-related negativity (ERN). Top: Response-synchronized waveforms for correct and error trials at FCz in an epoch running from 200 ms before until 500 ms after the response. The ordinate indicates the time of the response. Bottom: Scalp voltage maps at the time of the peak of the ERN (56 ms after the response), separately for correct trials, error trials, and the difference between these conditions.

difference between congruent and incongruent trials to measure the effect of response conflict, pairwise comparisons revealed that the effect of congruence was larger at FCz than at other electrode sites (although only marginally so at Cz, $p < .06$; $p < .01$ at all other electrodes).

Thus, both the ERN and N2 are clear in the data. We now turn to direct comparisons between these components. Our first prediction is that the ERN and N2 should share a similar topography and neural source. Our second prediction is that the N2 should precede the response whereas the ERN should follow it.

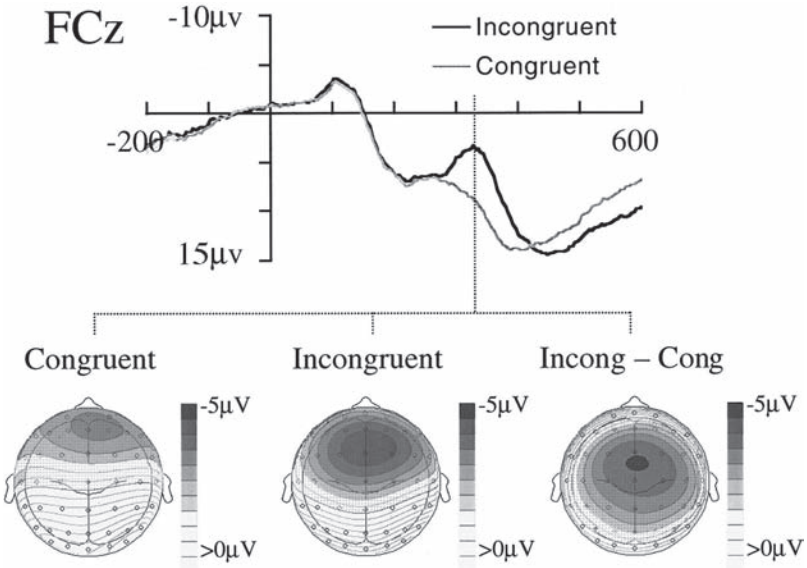

Figure 11. The N2. Top: Stimulus-locked waveforms for correct trials, separately for congruent and incongruent stimuli. Data are shown for electrode FCz, in an epoch running from 200 ms before until 600 ms after the stimulus. Bottom: Scalp topography for congruent and incongruent trials, and the difference between these conditions observed 344 ms after stimulus presentation. The scalp topography data for congruent and incongruent trials were high-pass filtered above 2 Hz to remove the contribution of slow parietal positivities that otherwise mask the effects of interest. Incong = incongruent; Cong = congruent.

Comparison of scalp topography and neural source. As described above, and as shown in Figures 10 and 11, the ERN and N2 share a similar scalp topography with peak amplitude at FCz. To compare the topographies more directly, we first scaled the data so that the ERN and N2 were equated for amplitude (McCarthy & Wood, 1985). To this end, we calculated the best fitting regression line between voltage amplitude for the ERN and N2 difference waves across electrodes. This analysis indicated that there was a high degree of similarity in topography between the two components, $r(63) = .80$, $p < .01$, and that ERN amplitude was 1.9 times as large as N2 amplitude. This value was used to equate the amplitude of the N2 and ERN when comparing their scalp distributions.

An ANOVA performed on the scaled data revealed that the N2 had a slightly more posterior and right-lateralized topography than the ERN, reflected in significant interactions between component and anterior-posterior location, $F(4, 60) = 11.9$, $MSE = 4.0$, $\varepsilon = 0.41$, $p < .01$, and between component and laterality, $F(2, 30) = 4.97$, $MSE = 3.95$, $\varepsilon = 0.90$, $p < .05$, although the three-way interaction was not reliable, $F(8, 120) = 1.84$, $p > .1$. Inspection of the scalp voltage maps (Figures 10 and 11), however, indicated that these topography differences were not consistent across conditions. In particular, the N2 observed on incongruent trials showed no evidence of a focus lying posterior to FCz, and little evidence of right-lateralization. Indeed, a comparison of the ERN and the incongruent trial N2, $r(63) = .93$, $p < .01$, scaling factor = 2.3, revealed a marginally reliable trend for a more *frontal* focus for the N2 compared with the ERN, $F(4, 60) = 2.79$, $MSE = 3.27$, $\varepsilon = 0.35$, $p < .1$.

A general problem when comparing scalp topographies across conditions is that any given voltage distribution is likely to be the result of many overlapping components, each reflecting the activity of a different neural source (Coles, Gratton, & Fabiani, 1990). One method for dealing with this problem is to compare dipole source models of the observed scalp voltage distributions that are stable to the addition of further dipoles to the solution (for a similar logic, see Miltner et al., 1997). We therefore performed such an analysis for the N2 and ERN. As one would expect given their similar topographies, the best fitting dipoles for the two components lay very close together in medial frontal cortex, as shown in Figure 12. The single dipole models explained most of the variance in the data for the ERN (dipole location: x = −0.4 cm, y = 1.1 cm, z = 5.7 cm; residual variance [RV] = 7.2%), and for the N2 difference wave (dipole location: x = −0.5 cm, y = 1.1 cm, z = 4.6 cm; RV = 6.8%).[5] Although the ERN and N2 dipole locations were not identical, the observed difference of 1.1 cm is well within the range of variability seen in estimates of the dipole source of the ERN across experimental conditions (e.g., Dehaene et al., 1994; Holroyd et al., 1998). Even greater variability is observed in the localization of the N2 across conditions of the same experiment (Lange et al., 1998; Liotti et al., 2000). In addition, the data from both conditions were well fit by a single dipole located halfway between the best fitting locations (ERN: RV = 7.2%; N2: RV = 8.4%). The present results are therefore consistent with the hypothesis that the ERN and N2 have a common neural source within medial frontal cortex.

Relative timing. The upper panel of Figure 13 plots response-synchronized waveforms for correct congruent, correct incongruent, and error trials at electrode FCz, with a baseline taken from

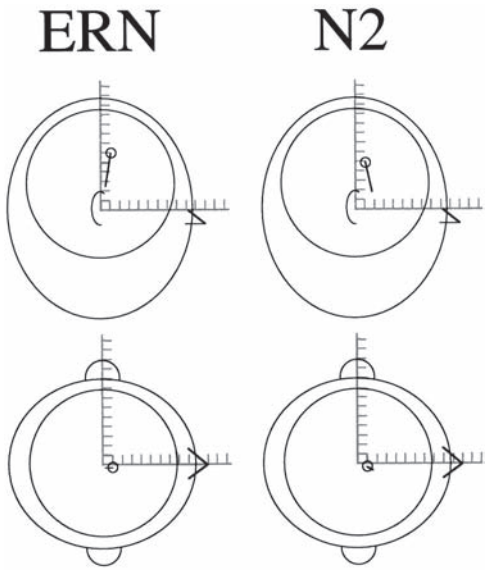

Figure 12. Dipole models of the sources of the error-related negativity (ERN) and N2 difference waves.

800 to 700 ms prior to the response. The data were high-pass filtered with a low cutoff of 2 Hz so that the frontocentral negativities of interest are not masked by large, slow positive waves apparent around the response. Figure 13 shows that there was a period, just prior to the response, of enhanced negativity on incongruent trials compared with congruent trials. To quantify this effect, we calculated the average voltage in a 100-ms window centered on –100 ms pre-response, separately for the 15 electrode locations used in the earlier analyses. This analysis revealed that the waveform on incongruent trials was significantly more negative than on congruent trials, $F(1, 15) = 5.11$, $MSE = 5.97$, $p < .05$. The amplitude of this congruence effect was largest at site FCz, indicated by reliable interactions between congruence and anterior-posterior location, $F(4, 60) = 35.3$, $MSE = 0.49$, $\varepsilon = 0.29$, $p < .01$, between congruence and laterality, $F(2, 30) = 7.77$, $MSE = 0.28$, $\varepsilon = 0.92$, $p < .01$, and between congruence, anterior-posterior location, and laterality, $F(8, 120) = 4.93$, $MSE = 0.02$, $\varepsilon = 0.37$, $p < .01$. Pairwise comparisons revealed that the effect of congruence at FCz was highly reliable ($p < .01$) and was larger than the effect of congruence at other electrode locations ($p < .01$ at all electrodes except Fz, at Fz $p = .20$).

The difference between congruent and incongruent trial waveforms was largest 88 ms prior to the response. The middle panel of Figure 13 shows the scalp topography of each waveform at this time. A frontocentral negativity was apparent on incongruent trials

[5] The coordinates of dipole locations are given in a reference frame based on the location of the nasion and preauricular points. The x value indicates laterality relative to the midpoint of an axis joining the two preauricular points (with negative values indicating points to the left of the midline). The y value gives the distance of the dipole from this axis toward the nasion (with positive points lying anterior to the preauricular line). The z value gives the distance of the dipole in the dorsal-ventral direction, orthogonal to the plane formed by the nasion and preauricular points (with positive points lying above this plane).

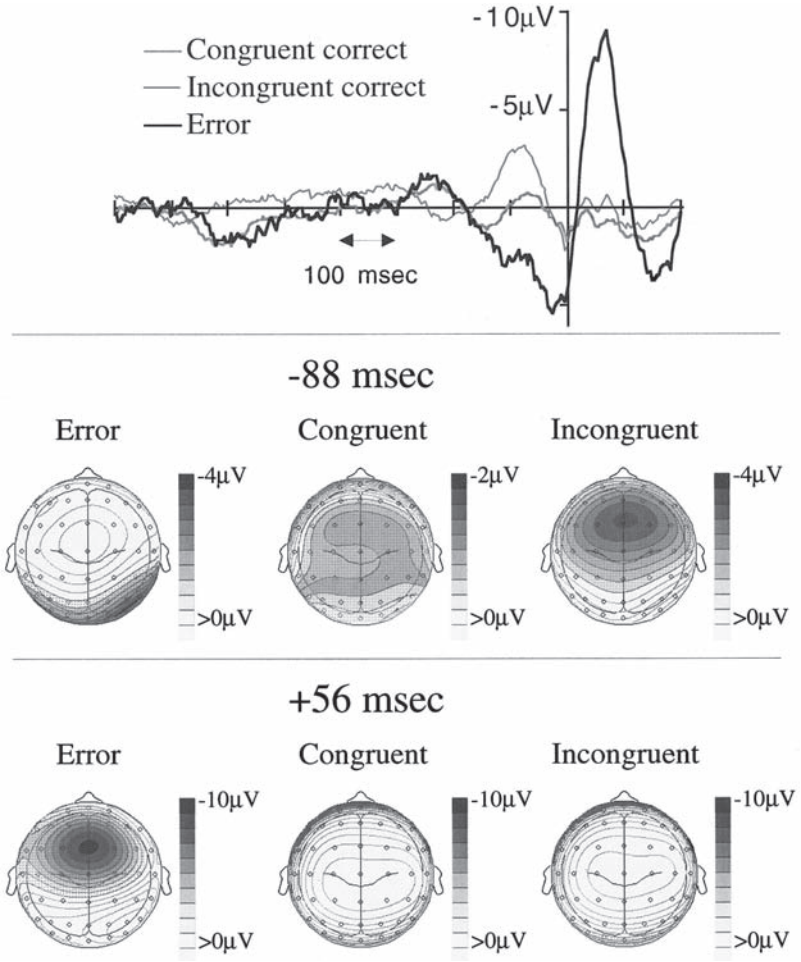

Figure 13. The relative timing of the N2 and error-related negativity. The upper panel shows response-synchronized waveforms at FCz for correct congruent, correct incongruent, and error trials. The ordinate indicates the time of the response. The data were high-pass filtered above 2 Hz to remove the contribution of slow parietal positivities. The other panels show scalp voltage maps for error trials, correct congruent trials, and correct incongruent trials at 88 ms before the response (middle panel) and 56 ms after it (lower panel).

and, to a lesser extent, on congruent trials also. Consistent with this being the N2 as observed in the stimulus-locked averages, the common dipole solution derived above gave a good fit to the scalp distribution observed on incongruent trials (RV = 10.9%).

The lower panel of Figure 13 shows voltage maps for each condition at the peak of the ERN, 56 ms after the response. The waveforms on congruent and incongruent correct trials differed little in the period following the response and showed no evidence of a negative component at the latency of the ERN (cf. Pailing et al., 2000; Scheffers & Coles, 2000; Ullsperger & von Cramon, 2001). In contrast, a frontocentral negativity was observed on error trials at this latency. The scalp topography of this negativity resembles closely that observed on incongruent trials 144 ms earlier.

Figure 14 shows a direct comparison between the predictions of the model and the empirical data at FCz. The simulated N2 is the difference between correct incongruent and correct congruent trials, the simulated ERN is the error-minus-correct difference wave. The model provides a good account of the relative timing of the two components. The difference in peak latency for the empirically observed N2 and ERN was 144 ms. The corresponding value in the simulation data was 9 cycles (≈144 ms), a strikingly good fit given that the model was not parameterized specifically to fit these data. The model gives a less accurate simulation of the relative amplitudes of the two components, however, failing to reproduce the large asymmetry seen in the data, and also overestimates the duration of the two components, discrepancies that might be addressed in future research.

Further properties of the N2. It might be objected that the appearance of the N2 in the response-locked waveforms was simply an artifact of averaging across trials in which a stimulus-locked N2 occurred just prior to the mean RT. To demonstrate that this is not the case, and to illustrate further properties of the N2, in Figure 15 (left panel) we present stimulus-locked waveforms for correct trials divided into sequential RT bins of 50 ms (cf. Ritter, Simson, Vaughan, & Friedman, 1979). An N2 is apparent in all but the fastest RT bins, with a latency and amplitude that vary systematically as a function of RT. Specifically, the N2 increased in

amplitude and latency with increasing RT. N2 amplitude was only slightly larger for incongruent than for congruent trials within each RT bin: The overall difference in N2 amplitude between the conditions evident in Figure 11 reflects the fact that congruent trials fell largely in the faster RT bins (with small N2 amplitude), whereas incongruent trials tended to have longer RTs (and large N2s).

Quantification of these properties was complicated by the fact that not all participants produced RTs in every 50 ms bin. We instead selected three consecutive RT bins for each participant so that there were appreciable numbers of trials, both congruent and incongruent, in each bin. For 9 of the participants, the bins ran from 350 to 500 ms, for 4 of the participants, the bins ran from 300 to 450 ms, and for 3 participants the bins covered 250 to 400 ms. For each bin—fast, medium, and slow—we calculated N2 amplitude and latency separately for the congruent and incongruent trial waveforms (that were first low-pass filtered below 20 Hz). N2 latency was defined as the time of the most negative peak in a window from 200 to 400 ms after the stimulus, and N2 amplitude was defined as the difference in voltage between this peak and the immediately preceding most positive peak. For 3 participants, no negative peak was apparent in the waveforms in one or more conditions. In these cases, the amplitude of the N2 was defined as zero. In addition, because the latency is undefined for conditions with no negative peak, we did not include the data from these 3 participants in the latency analysis.

N2 amplitude varied significantly as a function of RT, $F(2, 30) = 8.84$, $MSE = 11.0$, $\varepsilon = 0.74$, $p < .01$, with a significant linear trend, $F(1, 30) = 14.8$, $MSE = 162.1$, $p < .01$. Pairwise comparisons revealed that N2 amplitude was significantly ($p < .01$) larger for trials in the slow RT bin (-8.0 μV) than for trials in the medium (-5.2 μV) and fast (-4.8 μV) bins, which did not themselves differ. In addition, even though the conditions were matched for RT, N2 amplitude was somewhat larger for incongruent trials (-6.9 μV) than for congruent trials (-5.1 μV), $F(1, 15) = 12.3$, $MSE = 6.8$, $p < .01$.[6]

N2 latency increased with RT, $F(2, 24) = 5.2$, $MSE = 685.2$, $\varepsilon = 0.73$, $p < .05$, with a reliable linear trend, $F(1, 24) = 7.0$, $MSE = 4{,}807.7$, $p < .05$. Pairwise comparisons revealed that the N2 peaked significantly later on slow trials (320 ms) than on medium RT (299 ms, $p < .01$) or fast RT (300 ms, $p < .05$) trials, which did not differ. N2 latency was also slightly increased for incongruent trials (312 ms) relative to congruent trials (301 ms), again a reliable difference, $F(1, 12) = 6.7$, $MSE = 356.6$, $p < .05$. Importantly, interparticipant variability in N2 latency did not in-

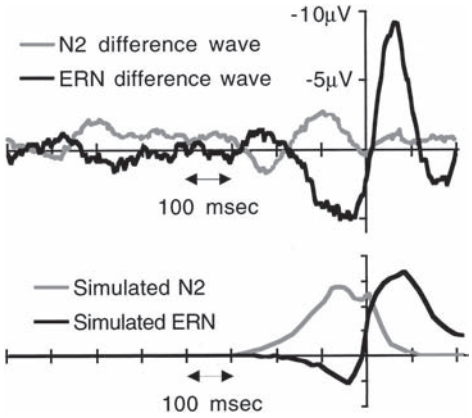

Figure 14. Comparison of observed (top) and predicted (bottom) latencies of the N2 and the error-related negativity (ERN).

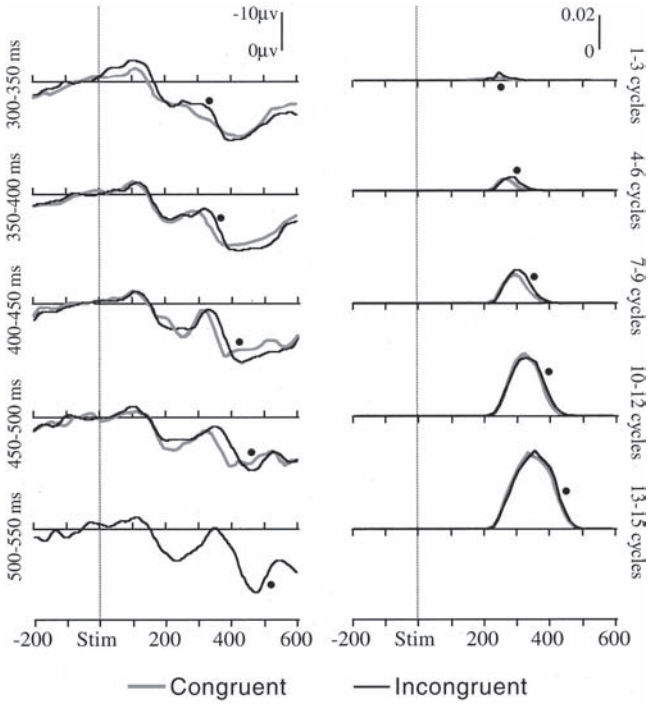

Figure 15. The left column shows event-related brain potential (ERP) waveforms for correct congruent and incongruent trials, separated into sequential reaction time (RT) bins from 300–350 ms (top panel) through to 500–550 ms (bottom panel). The dotted line indicates the time of the stimulus (Stim), and the average RT for each bin is indicated with a circle. The right column shows plots of simulated response conflict, with trials separated into sequential RT bins of 3 cycles (~48 ms). Note that there were too few trials to obtain reliable ERPs for congruent trials in the 500–550 ms bin. In addition, 4 participants produced no responses on incongruent trials in the 300–350 ms bin, and 2 participants produced no responses in the 450–500 ms bin on congruent trials: The waveforms for these conditions are based on the data of the remaining participants.

[6] An analysis excluding the 3 participants for whom no N2 peak was evident in one or more of their ERP waveforms found essentially identical results to those reported above. Thus, the reported effects are not an artifact of base-to-peak measure that we used, in which N2 amplitude was set to zero when there was no peak in the ERP waveform. It is also important to demonstrate that the apparent increase in N2 amplitude with RT is not an artifact of differential overlap with the P3 component across RT bins. If the observed effects of RT reflected changes in the P3 component, then such effects should be even more marked at posterior scalp sites where the P3 is maximal. However, this was not the case: Voltage differences between RT bins at the peak latency of the N2 were larger at FCz than at parietal site POz, $F(2, 30) = 4.4$, $MSE = 8.5$, $\varepsilon = 0.71$, $p < .05$. We also conducted a temporal principal-components analysis (PCA) with varimax rotation to separate the contributions of the N2 and P3 to the ERP. The PCA was conducted on averaged ERP waveforms at electrodes FCz and POz, with separate waveforms for each participant, for congruent and incongruent

crease with RT (fast RTs, $SD = 39$ ms; medium RTs, $SD = 30$ ms; slow RTs, $SD = 25$ ms). Thus, the broadening of the N2 peak with increasing RT was not an artifact of increased variability in peak latency, but instead reflected a real increase in the amplitude of the N2 component with increasing RT.

In a third analysis, we calculated the latency difference between the peak of the N2 and the average RT, separately for each RT bin. This analysis revealed that the N2 peak occurred reliably earlier than the key-press response (mean latency difference = 87 ms), $F(1, 12) = 65.1$, $MSE = 4{,}530.1$, $p < .01$. The interval between the N2 peak and the response increased reliably with RT, $F(2, 24) = 44.5$, $MSE = 746.6$, $\varepsilon = 0.72$, $p < .01$, with a significant linear trend, $F(1, 24) = 87.2$, $MSE = 65{,}107.7$, $p < .01$: The N2–RT interval was larger for trials in the slow RT bin (120 ms) than in the medium bin (93 ms), and was smallest in the fast RT bin (49 ms). These latency differences across RT bins were all reliable ($p < .01$). The N2–RT interval did not differ between congruent and incongruent trials, $F(1, 24) = 1$.

To determine whether the conflict monitoring theory can account for these detailed properties of the N2, we reanalyzed the results of Simulation 1, calculating response conflict on correct trials in sequential RT bins of 3 cycles (~48 ms). The results of this analysis, shown in the right-hand panel of Figure 15, accurately replicate a number of features of the empirically observed N2. First, the conflict signal, like the N2, becomes larger and broader as RT increases. This feature is present in the model because conflict between the two response units delays production of the response, so that RT increases with the degree of conflict observed. A second feature of the simulation results is that response conflict is only slightly larger on incongruent trials than on congruent trials when the conditions are matched for RT: In the model, slow responses to congruent stimuli are marked by high conflict (because of noise in processing) in just the same way as are slow responses to incongruent stimuli. The overall difference in conflict between congruent and incongruent trials reflects the fact that a greater proportion of congruent trials fall in the faster RT bins (as a result of low conflict), whereas incongruent trials tend to have greater RTs (as a result of high conflict). Finally, the interval between the peak of the conflict signal and the response increases with RT. This property of the model replicates the observation that the N2–RT interval increases with RT. The present theory provides a simple explanation of this finding: The greater the conflict at any given time, the longer it should take to resolve this conflict and execute one of the competing responses.

Thus, the simulated data replicate quite closely the patterns observed in the empirical data, and particularly so in the faster RT bins. In the 450–500-ms bin, the relationship between empirical and simulated data is less good: Although empirical N2 amplitude is similar for congruent and incongruent trials when measured base to peak, as is the simulated conflict signal, the latency of the empirical N2 is increased on incongruent trials, a finding not seen in the simulation data. This discrepancy is worthy of note, but should be interpreted with caution because there were very few trials in the empirical data for each participant in the slower RT bins, particularly for the congruent condition. In the faster RT bins, for which there were many more trials per condition and hence for which the ERP waveforms were more stable, the patterns in the empirical N2 data are very similar to those predicted by the conflict theory.

Discussion of ERP Results

On the basis of the simulation results, we made two predictions that were tested in the present experiment. These predictions concerned the relationship between the ERN and the N2, a component that we hypothesize to be a correlate of conflict monitoring on trials with correct responses. Both of the predictions were supported by the data. First, the ERN and N2 shared a very similar scalp topography and neural source (converging results have recently been reported independently by van Veen & Carter, 2002). Second, the two components differed in their timing, with the N2 preceding the response and the ERN following it. A further comparison between the N2 and the simulated conflict signal revealed that detailed properties of the N2 can be explained by the conflict monitoring theory.

Conflict, errors, and ACC function. The present findings may help to interpret findings from fMRI studies of ACC function by Kiehl et al. (2000) and Menon et al. (2001). These studies found a common region in caudal ACC that was activated on error trials and on correct trials with conflict. However, they also found a region of rostral ACC that activated only on error trials. Both Kiehl et al. and Menon et al. linked this rostral ACC activation with the ERN. However, our results, together with converging findings from van Veen and Carter (2002), suggest that the ERN is generated in a region of ACC that is sensitive to conflict, most likely caudal ACC. It is possible that rostral ACC activation following errors reflects further, perhaps affective, processing of the error, because this part of ACC is thought to be associated with affective function (Bush et al., 2000; Devinsky, Morrell, & Vogt, 1995). Van Veen and Carter have provided evidence in favor of this hypothesis: Their dipole models of the Pe included a source in rostral ACC that became active some 200 ms after the initial error.

Another issue to address concerns a recent report from Davies, Segalowitz, Dywan, and Pailing (2001) of a dissociation between the ERN and N2. They reasoned that if the ERN and N2 are related, then participants with a large ERN should also show a large N2. Davies et al. performed separate correlations between ERN amplitude and N2 amplitude on congruent trials and between the ERN and N2 amplitude on incongruent trials and found no significant correlations. However, of more relevance to the present hypothesis is the relationship between the ERN and the N2 difference wave—that is, the difference between incongruent and congruent trials. This difference measure would seem to be a purer measure of response conflict on correct trials. When we analyzed our data in this way, we found a reliable correlation between the

trials, and for the fast, medium, and slow RT bins (a total of 192 ERPs). The first two components revealed by the PCA corresponded to the P3 and N2, respectively. Critically, the weighting of the N2 component varied significantly as a function of RT, $F(2, 30) = 14.9$, $MSE = 6{,}260.3$, $\varepsilon = 0.65$, $p < .01$, and more so at FCz than at POz, $F(2, 30) = 5.08$, $MSE = 1{,}588.3$, $\varepsilon = 0.74$, $p < .05$. Pairwise comparisons revealed that N2 amplitude at FCz was larger for slow trials than for medium trials ($p < .01$) and was larger for medium trials than for fast trials ($p < .01$), consistent with the results of our analysis of the raw ERP waveforms. Thus, we conclude that the effects of RT on N2 amplitude reflect real changes in the N2 component, not changes in the overlap between N2 and P3.

amplitudes of the ERN and N2 difference wave across subjects, $r(15) = .60$, $p < .05$. That is, participants with a greater sensitivity to conflict showed a corresponding greater sensitivity to errors. Although it may be possible to generate alternative accounts of this finding, the results are certainly consistent with our theory that both components reflect a common function, conflict monitoring.

Relation to existing theories of N2 function. We hypothesize that the N2 component observed in the flanker task may be a correlate of conflict monitoring. This hypothesis leads us to predict that conflict-related N2 components should be apparent in other situations characterized by high response conflict. It is therefore of interest that N2 components are apparent in the oddball (Ritter et al., 1979; Ritter, Simson, Vaughan, & Macht, 1982) and go–nogo tasks (Kok, 1986; Pfefferbaum, Ford, Weller, & Kopell, 1985). In the oddball task, participants respond to infrequent targets and withhold responses to frequent distractors; in the go–nogo task, participants are required to withhold the prepotent go response to a subset of the stimuli (nogo condition). As noted by Braver et al. (2001), both oddball and go–nogo tasks should produce high response conflict, because both require participants to overcome a prepotent response tendency. Thus, we suggest that N2 components observed in the oddball and go–nogo tasks may have a common origin in conflict monitoring by ACC.

The conflict monitoring theory is broadly consistent with the proposal of Ritter and colleagues (Ritter et al., 1979, 1982) that the oddball N2 is related to decision or categorization processes. The present theory extends this earlier account by specifying precisely which aspect of the decision process—response conflict—is reflected in the N2. In this way, the conflict monitoring theory provides a natural account of previous findings relating the N2 to response selection. For example, Ritter et al. (1979) have shown that the oddball N2 peaks around 100 ms prior to the response and that its amplitude is increased on trials with longer RTs, consistent with our simulation and empirical data from the flanker task.

Our theory contrasts with previous accounts of the nogo N2, which typically associate this component with response inhibition (e.g., Kok, 1986; see also Kopp et al., 1996). Nieuwenhuis, Yeung, van den Wildenberg, and Ridderinkhof (2003) have recently compared the inhibition and conflict monitoring accounts of the nogo N2. They reasoned that if the nogo N2 reflects response conflict, then it should have a similar neural source and timing to the oddball N2. Following up the present findings, they also predicted that the oddball and nogo N2 components would share a neural source with the ERN. The results were consistent with these predictions. Thus, the conflict monitoring theory not only provides an integrative account of the ERN and N2, but also provides a unified account of N2 components observed in a variety of experimental tasks.

3. Error Detection Through Response Conflict

The previous sections have outlined in detail our hypothesis that the ERN can be explained in terms of the conflict monitoring theory. According to this hypothesis, the ERN is not an explicit signal that an error has occurred, but is rather a signal that there is response conflict. In this regard, our theory contrasts with existing accounts of the ERN that associate this component with an explicit error detection process (e.g., Coles et al., 2001; Falkenstein et al., 1991, 1995, 2000; Gehring et al., 1993). However, in seeking to explain the ERN in terms of conflict monitoring rather than explicit error detection, our theory appears to leave open the question of how people are able to detect their errors. To answer this question, we now introduce a new theory of how errors may be detected in the brain. The basis for this theory is the observation in our simulations that error trials are characterized by conflict that develops in the period following the response. Given that this post-error conflict replicates many properties of the ERN, and given that the ERN demonstrates many properties expected of an error-detection system, it seems possible that monitoring for response conflict might represent a simple method for detecting errors. To evaluate this hypothesis, we next present an analysis of the performance of a simple system that detects errors on the basis of the total amount of conflict observed in the period following the response. We then compare the performance of this system with empirical findings about human error-detection performance.

Performance of a Conflict-Based Error Detector

Consider a system that signals an error has occurred whenever the amount of conflict in the post-response period exceeds some threshold. On some proportion of error trials, continued stimulus processing will lead to post-response activation of the correct response, resulting in enough post-response conflict to exceed the detection threshold. This will result in the system correctly detecting the error—these are *hit* trials. On other error trials, however, the incorrect response will continue to dominate even after the response has been produced. The conflict signal would remain below threshold, and the monitoring system would incorrectly signal the trial to be correct—these are *miss* trials. The proportion of error trials for which the monitor signals an error is given by $P(d|error) = \Sigma \text{ hits} / (\Sigma \text{ hits} + \Sigma \text{ misses})$, where *d* stands for "detection of an error signaled"—that is, a threshold crossing in the post-response conflict signal. The sum, Σ, is simply the number of the relevant events in the experiment. Similarly, we can calculate $P(d|correct) = \Sigma \text{ FA} / (\Sigma \text{ FA} + \Sigma \text{ CR})$, where FA stands for false alarm and CR stands for correct rejection. False alarms would occur when activation of the incorrect response unit followed a correct response, leading to a suprathreshold conflict signal following the response on a correct trial. Correct rejections would occur under more normal circumstances, where continued stimulus processing simply reinforced the correct response decision, and little or no conflict followed the response.

It is possible to assess the performance of the model by comparing P(d|error) and P(d|correct) for a given detection threshold. Good performance is indicated by a high value of P(d|error), indicating that most errors are detected, accompanied by a low value of P(d|correct), indicating few false alarms. Plotting P(d|error) against P(d|correct) for a range of detection thresholds gives the receiver operating characteristic (ROC) curve for the model. Figure 16 (left panel) plots ROC curves for the three speed–accuracy conditions of Simulation 3. This analysis shows that error detection based on post-response conflict is most reliable in the accuracy condition and is least reliable in the speed condition. Detection in the accuracy condition appears to be quite reasonable. The point marked on the graph, for example, indicates that a conflict monitor could detect 75% of all errors, while giving false alarms on just 4.5% of correct response trials.

However, given the very different base rates of errors and correct responses, this detection performance would represent a very unreliable error signal. In 1,000 trials of the accuracy condition, for example, the model made 910 correct responses and only 90 errors. Thus, of the error signals produced by the model using this detection threshold, 41 ($\approx$ 910 * 0.045) would be false alarms and only 68 ($\approx$ 90 * 0.75) would be hits. Thus, of the 109 "error signals" produced by the model, only 62% would accurately signal an error. To illustrate this point, the right-hand panel of Figure 16 gives P(d|error) as a function of P(error|d) for a range of conflict threshold values, where P(error|d) = Σ hits / (Σ hits + Σ FA), the probability that a threshold crossing correctly signals an error. Good error-detection performance is indicated by points toward the top right of the graph (i.e., most errors being detected and most threshold crossings corresponding to errors). According to this analysis, the post-response conflict signal is a poor error signal, with no points lying in the top right-hand corner of the graph for any of the three speed–accuracy simulations. With a simple modification, however, error-detection performance can be dramatically improved.

The reason for the poor performance described above is that, even for correct trials, conflict is not completely resolved by the time of the response, but rather continues to be present for a few cycles afterward. It is therefore possible to improve the sensitivity of the error monitor by measuring conflict only after some delay following the response, by which time the correct-trial conflict is completely resolved and any persisting conflict more specifically signals an error. Figure 17 presents a plot of P(d|error) as a function of P(error|d) for the neutral condition of Simulation 3, separately for analyses in which the measurement of post-response conflict began 0, 2, 4, or 6 cycles after the response. The sensitivity of the model is greatly improved by introducing a delay in this manner. In the delay 6 condition, for example, the point marked x on the graph indicates that a suitably thresholded conflict monitor could detect 92% of errors, with over 80% of threshold crossings corresponding to errors. Recall that this detection performance holds in a condition with a high error rate (15%) and in which only 63% of errors were corrected. The corresponding calculation for the accuracy simulation produced an optimal detection rate of 95%

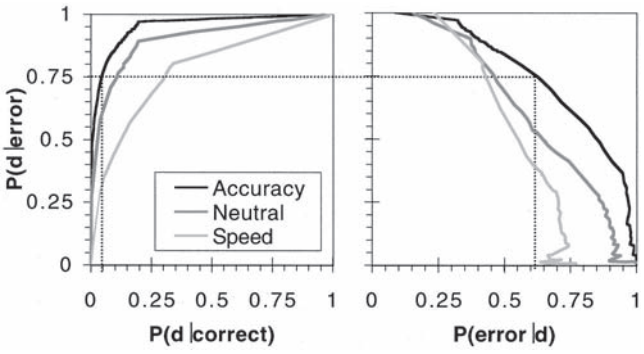

Figure 16. Performance of a conflict-based error-detection mechanism applied to data from the three speed–accuracy conditions of Simulation 3. Left: The probability of error detection as a function of false alarm proportion. Right: Error-detection rate as a function of the proportion of threshold crossings that correspond to errors. The dotted line shows performance for one threshold value in the accuracy condition.

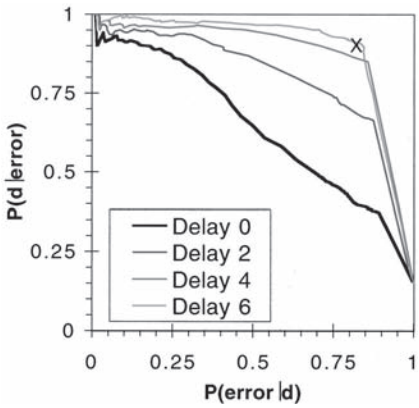

Figure 17. Performance of the conflict-based error detector with delays of 0, 2, 4 and 6 cycles between the response and the conflict monitoring window. The probability that errors are detected is plotted as a function of the probability that detection signals correspond to errors.

of errors, with less than 3% of all threshold crossings being false alarms. This excellent performance was obtained in a condition in which errors were still quite frequent (error rate = 9.9%) and in which 86% of errors were corrected. Thus, monitoring for post-response conflict could in principle be a reliable method for detecting errors. We next relate the results of this analysis to existing data and theory concerning human error detection.

Comparison With Human Error Detection

Rabbitt (1967, 1968, 2002) has conducted a number of studies of participants' error-detection performance. In each of his experiments, participants made errors on around 5% of the trials and were able to detect 80%–95% of those errors given sufficient time. The observed rate of error detection is comparable to the performance of our conflict-based error detector, particularly for the accuracy condition of Simulation 3 in which error rates were similarly low (though still 9.9%). In addition, Rabbitt's participants made occasional false alarms, signaling that they had made an error when in fact they had not. These false alarms constituted 7%–9% of all signaled errors, again comparable to the performance of our simple conflict-based error detector. Overall, the quantitative performance of the model is comparable to that of human participants.

The present theory is also consistent with findings from Rabbitt's (1967, 1968, 2002) experiments concerning the relationship between error detection and error correction. The model implements the idea that continued processing of the stimulus following errors leads to activation of the correct response. That is, there is a natural tendency for the model to correct its own errors, a tendency that forms the basis for our theory of the ERN. In contrast to automatic error correction, our theory holds that explicit error detection involves the computation of conflict for some period after an error is committed and, hence, will be a slower process. This property of the model is consistent with Rabbitt's (1968, 2002) finding that participants respond to errors more quickly and efficiently with a correcting response than when making a common detection response to all errors.

As further evidence of the fast, automatic nature of error correction, Rabbitt et al. (1978) highlighted the very fast error-correcting responses that are often observed in RT experiments. Such responses can be observed to follow within 10–20 ms of the error. Rabbitt et al. noted that it is unlikely that participants could detect and then correct errors within such short intervals. Instead, the finding seems more consistent with the idea that participants sometimes initiate two separate responses in close succession. Of interest in this regard, the model also corrects some errors very quickly—within 1–3 cycles of the initial error—and does so, of course, without having to detect the initial error. Thus, the model is consistent with Rabbitt's intuitions on two counts: First, error correction may occur automatically in a system in which information flow is continuous and increasingly accurate over time; and, second, error correction may in some sense precede error detection.

A prediction of our theory is that participants' ability to detect their own errors should depend systematically on the experimental context. In the preceding analysis, for example, different speed–accuracy conditions yielded distinguishable ROC curves for error detection. We therefore predict that participants encouraged to be more accurate will not only make fewer errors but will also more accurately detect those errors that they make. Some preliminary support for this prediction is provided by a reanalysis of the error-detection data reported by Rabbitt (1967). Across the 12 conditions of the experiment, there was a reliable negative correlation between error rate and rate of error detection, $r(11) = .63$, $p < .05$, consistent with our prediction. Although it may be possible to derive alternative explanations of this finding, the results are at least in line with the predictions of our theory. Future experiments could provide a more rigorous, quantitative test of this prediction.

Discussion of Conflict-Based Error Detection

Taken together, our analyses demonstrate that monitoring for conflict in the period following a response could serve as a method for detecting errors. This is the case even though the underlying process does not involve a direct evaluation of response accuracy, nor does the process involve an explicit comparator mechanism. Instead, it may be possible to detect errors on the basis of a feature of processing—the occurrence of conflict following a response—that is statistically associated with incorrect responses. Although this account of error detection suggests additional mechanisms beyond the conflict monitoring unit simulated in our model, the required computation is very simple: Our analysis suggests that adequate error detection can be performed using a straightforward accumulator that signals an error has occurred whenever post-response conflict exceeds a threshold.

The present theory shares with previous accounts the hypothesis that error detection relies on the fact that continued stimulus processing will tend to produce an increasingly reliable representation of the correct response within the task-processing system responsible for the initial error (Falkenstein et al., 1991, 2000; Rabbitt et al., 1978; Rabbitt & Vyas, 1981; Scheffers and Coles, 2000). However, our theory suggests a different view of the relationship between the ERN and error detection than has been proposed in previous theories. In particular, according to our theory, the ERN does not index the process of error detection itself, but rather reflects post-response conflict that develops following errors. However, as the present analysis demonstrates, this hypothesis should not be taken to imply that the ERN has nothing at all to do with error detection: Even if the ERN is generated by conflict monitoring, it might nonetheless serve as an effective error signal with little additional machinery. Thus, whereas existing theories view the ERN as reflecting the output of an error detection process, our theory suggests that the ERN may in fact reflect the input to this process. Of course, the present demonstration does not constitute proof that human error detection in fact relies on conflict monitoring. Thus, an important goal for future research is to look for more direct evidence that error detection involves monitoring for response conflict. A critical step toward this goal will be to implement contrasting theories of error detection in comparable detail to ours, allowing predictions of the theories to be contrasted in a formal, quantitative manner. Our ongoing research has begun to address this issue (Holroyd, Yeung, Coles, & Cohen, 2004).

Another issue for future research is the extent to which conflict monitoring may be used to detect errors in other processing systems. The proposed mechanism for detecting errors makes use of a simple and quite general property of human information processing: that representations tend to become increasingly accurate over time. Hence, conflict monitoring may provide reliable information about errors in processing in a wide range of processing domains. For example, there are interesting parallels between the issues considered in this article and issues discussed in the literature on speech errors (Postma, 2000). Here again there is debate about whether error detection requires a dedicated monitoring system (e.g., Levelt, 1989) or whether mistakes can be detected on the basis of the processing dynamics that characterize speech errors (e.g., MacKay, 1992). We naturally favor the latter view and speculate that conflict between representations of intended and actual speech may be a reliable method for detecting speech errors.

General Discussion

The principal contributions of the present research may be summarized as follows:

A new theory of the ERN. According to our theory, the ERN reflects conflict that develops in the period following errors as a consequence of continued processing of the stimulus. Continued processing leads to post-error activation of the correct response and hence conflict with the incorrect response just produced. The simulation results demonstrate the ability of this theory to explain findings that have previously been interpreted as challenging it (Pailing et al., 2000; Scheffers & Coles, 2000; Ullsperger & von Cramon, 2001), to provide new accounts of existing findings (Gehring et al., 1993; Holroyd & Coles, 2002), and to provide insight into the cause of apparently discrepant empirical results (Gehring et al., 1993; Scheffers et al., 1996).

A new theory of the N2. We propose that the N2, like the ERN, is a correlate of conflict monitoring. Specifically, we suggest that the N2 reflects conflict in the period prior to the response on trials with correct responses. In this way, the conflict monitoring theory provides a unified account of the N2 components observed in the flanker, oddball, and go–nogo tasks (Nieuwenhuis et al., 2003) and provides an integrative account of the N2 and the ERN in terms of a common underlying mechanism.

A new theory of error monitoring. An analysis of the simulated dynamics of response conflict suggests that errors may be detected reliably by monitoring for conflict in the period following the response. The quantitative performance of a simple conflict-based error detector was comparable to that observed for human participants in empirical studies (Rabbitt, 1967, 1968, 2002). We therefore suggest that the brain may use conflict monitoring as a computationally simple method for detecting errors.

Comparison With Existing Theories

We have introduced a new theory of the ERN and error detection in terms of the conflict monitoring theory of ACC function. According to our theory, the ERN is not an explicit signal that an error has occurred, but rather reflects the continuous evaluation of response conflict that may, with simple additional mechanisms, be used to detect errors reliably. In this regard, our theory stands as an alternative to the view that error detection involves an explicit comparison—between the executed response and a separate representation of the correct or intended response—and that the ERN reflects a mismatch signaled during this process. Our theory also contrasts with a recent computational model of the ERN proposed by Holroyd and Coles (2002), in which the ERN is held to be a reinforcement learning signal conveyed to ACC. In what follows, we directly compare these theories with our own. We then discuss the issue of whether the ERN reflects directly the process of error detection or, rather, is an emotional reaction to errors, as has recently been suggested.

The ERN as a mismatch signal. The most common interpretation of the ERN is that it reflects the outcome of a comparator process that detects errors as mismatches between the actual response and knowledge about the correct or intended response (Coles et al., 2001; Falkenstein et al., 1991, 2000; Gehring et al., 1993; Scheffers et al., 1996; Scheffers & Coles, 2000). The representation of the actual response is presumed to rely on efference copy, whereas the representation of the correct or intended response is held to be derived from continued processing of the stimulus after the incorrect response is produced: The notion is that errors occur when response execution occurs impulsively, before the stimulus is fully processed and hence before the response selection system has settled on a final representation of the correct response. It was initially proposed that the error detector waited until this final outcome of the response selection process before making the comparison (Falkenstein et al., 1991), consistent with the intuitions of Rabbitt and colleagues (Rabbitt & Rodgers, 1977; Rabbitt & Vyas, 1981). However, the latency of the ERN appears to be relatively invariant with respect to the response (Leuthold & Sommer, 1999; Rodríguez-Fornells et al., 2002), leading to the suggestion that the mismatch process may be triggered by response execution itself (Coles et al., 2001).

It is difficult to compare our theory directly with the mismatch hypothesis, because this latter account has been described in different ways by different researchers and has yet to be formalized in computational terms, making it difficult to draw precise predictions from the theory. Nevertheless, there appear to be important similarities between the mismatch hypothesis and the theory proposed here. In particular, both accounts propose that errors are detected on the basis of continued stimulus processing to provide a representation of the correct response that conflicts, or is compared, with a representation of the incorrect response. Moreover, the conflict-based error detection mechanism we have proposed is sensitive to conflict only after the response and is in this sense triggered by response execution, just as is the mismatch detector in Coles et al.'s (2001) account. Therefore, to the extent that the conflict and mismatch accounts are part of the same class of theories of performance monitoring (those that rely on continued processing of the stimulus to detect errors), the present simulations make a broad contribution: They represent the first attempt to provide a formal analysis of what information relevant to error detection might be present in the response system and of when this information might become available to a monitoring system.

However, the existence of similarities between the conflict and mismatch accounts should not be allowed to obscure the fact that there are also important differences between them. In particular, whereas the mismatch hypothesis proposes that the ERN reflects the output of a system specifically devoted to error detection, the present theory associates the ERN with a process—conflict monitoring—that also occurs on correct trials and which may represent the input to, rather than the output from, the error detection system. Thus, there are critical differences in the properties and predictions of the conflict and mismatch theories. In what follows, we discuss the implications of two specific differences between the theories.

A first critical difference is that our theory explains the ERN in terms of the continuous evaluation of response conflict, whereas the mismatch hypothesis proposes that the ERN indexes the operation of a discrete error-detection process. The proposal that response conflict is monitored continuously allows our theory to provide a unified account of the ERN and N2, thus explaining the close relationship between these components that has been observed in nogo and oddball tasks (Nieuwenhuis et al., 2003) as well as in the flanker task (as in the present research, and by van Veen & Carter, 2002). In contrast, mismatch detection is held to be a discrete process that occurs only at the time of response execution (Coles et al., 2001) or at some point thereafter (Falkenstein et al., 1991), raising the issue of whether this hypothesis can account for the N2. Falkenstein and colleagues have proposed that mismatch detection occurs only at the end of the response selection process, and so this account does not seem able to explain the N2 (which occurs prior to the response). Coles et al.'s (2001) more recent hypothesis may be able to explain the occurrence of the N2 if one assumes that the mismatch process is not only triggered by response execution but also can be triggered by subthreshold response activation, *partial errors*, occurring on trials with high conflict. However, the notion of partial errors is not part of Coles et al.'s theory as currently specified, suggesting the need for this account to specify more precisely the conditions under which the mismatch process is triggered.

A second difference between the theories is that response conflict occurs whenever there is coactivation of competing responses, whereas a mismatch signal is held to be generated only when there is an explicit representation of the correct or intended response within the response selection system that differs from the representation of the executed response derived from efference copy (e.g., Scheffers and Coles, 2000). This property implies that, for a mismatch to be detected and an ERN generated, at the time of detection there must be more activation of the correct response than the incorrect response in the response selection system; otherwise, the intended response, according to the response selection

system, is still the incorrect one, and there should be no mismatch with efference copy of the executed response. However, this property makes it difficult for the mismatch hypothesis to account for the timing of the ERN: In our simulations, ERN onset is observed at a time when the incorrect response unit continues to be more active than the correct response unit (cf. Figure 2). Rodríguez-Fornells et al. (2002) provided electrophysiological evidence consistent with this aspect of the model: Their LRP data suggest that the correct response becomes more active than the incorrect response well after error commission, whereas the onset of the ERN precedes error commission. Hence, at the time of the ERN, the state of the response selection system—with greater activation of the incorrect response—should not mismatch with the efference copy representation of the executed (incorrect) response, apparently inconsistent with the mismatch hypothesis. In contrast, our simulations suggest that a small degree of correct response activation may produce sufficient conflict immediately preceding errors to begin generating an ERN, even if this activation remains below that of the incorrect response, and can therefore account for the timing of the ERN.

The present research thus raises a number of challenges for the mismatch hypothesis in accounting for detailed properties of the ERN and its relationship to the N2. It is possible that a formal instantiation of this hypothesis will account for the empirical data as well as does the conflict theory. Alternatively, the mismatch hypothesis may in the future be revised to address the challenges we have raised. For example, one might propose that the mismatch process is continuous in nature or can be triggered by partial errors, perhaps accounting for the N2. Additionally, one might propose that the mismatch process does not require an explicit representation of the correct response to detect errors, thus accounting for the early onset of the ERN. However, these changes would represent significant departures from the mismatch hypothesis as currently specified and would move this account closer to ours. In this regard, an important avenue for future research will be to attempt the kind of formal investigation of the properties and predictions of the mismatch hypothesis that we have provided here for the conflict monitoring theory. This endeavor will help to provide detailed answers to the questions raised above—regarding how the mismatch process is triggered, on what representations it depends, and when this information might become available—so that this theory can generate precise predictions that can be compared with those of our theory.

The ERN as a reinforcement learning signal. Holroyd and Coles (2002) have recently proposed another alternative to the mismatch account of the ERN. They suggested that the ERN reflects a reinforcement learning signal that is transmitted to ACC from the basal ganglia via the mesencephalic dopamine system. According to this theory, ACC does not itself monitor errors, but rather receives a signal from the basal ganglia indicating that outcomes of actions are better or worse than expected. The role of ACC is then to use this learning signal to adapt the response selection process. Thus, this reinforcement learning theory of the ERN formally instantiates the notion, first proposed in association with the mismatch hypothesis (Coles et al., 1998, 2001; Miltner et al., 1997), that the ERN reflects the operation of generic error processing system. However, a different mechanism of error detection is proposed: Response errors are held to be detected as conjunctions of stimuli and responses associated with negative outcomes, rather than as mismatches between actual and intended responses. Holroyd and Coles have implemented this idea using a simple neural-network architecture based on the method of temporal differences (Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1998) and used it to model the ERN in a reinforcement learning task and in the modified flanker task described in Simulation 4.

Support for the reinforcement learning theory comes from the finding that negative feedback elicits a negative ERP component that, like the ERN, has a frontocentral scalp topography and a neural generator in the region of ACC (Miltner et al., 1997). The co-localization of this *feedback ERN* and the ERN observed immediately following error commission is consistent with the claim of the reinforcement learning theory that the ERN reflects the operation of a generic error processing system. In contrast, the conflict monitoring theory cannot presently account for the observation of the feedback ERN, which may be observed even in the absence of overt responses (Yeung, Holroyd, & Cohen, in press). At the same time, however, the reinforcement learning theory does not predict the observation of ACC activity on trials with correct responses and, hence, cannot account for the co-localization of the ERN and N2 that was observed in the present research (and also by Nieuwenhuis et al., 2003; van Veen & Carter, 2002). Thus, although both theories are currently supported by evidence from dipole modeling suggesting that the ERN co-localizes with a second component thought to provide an additional index of the functioning of ACC, the identity of this second component differs for the two theories, a discrepancy that remains for future research to resolve.

Further work is therefore required to distinguish between or reconcile the conflict monitoring and reinforcement learning accounts of the ERN, N2, and feedback ERN. In this regard, it is important to note that although the theories are superficially very different, they are not necessarily mutually exclusive. In particular, whereas our research has focused on how response errors might be detected on-line, Holroyd and Coles's (2002) model is primarily concerned with the issue of how information about response errors is integrated with other evaluative information and then used in response selection. Indeed, in their simulation of the data from the modified flanker task (cf. Simulation 4, above), Holroyd and Coles did not model any explicit error detection process; instead, information about response accuracy was simply provided to the model as an external signal. Thus, it might be that errors are detected through conflict monitoring, as we suggest, but that this information guides response selection through the kind of reinforcement learning framework envisioned by Holroyd and Coles. This integration of the conflict monitoring and reinforcement learning models of the ERN would help to address outstanding issues with each approach: For the response conflict theory, the reinforcement learning framework provides a way of understanding how a conflict-based error signal might be integrated with other evaluative information—such as performance feedback—in the learning process and the adaptation of behavior. For the reinforcement learning theory, the conflict framework provides a computationally simple method for generating reliable error signals and also provides a way of understanding why ACC should be so consistently activated in conditions of response conflict, even in the absence of overt response errors (as evident in fMRI studies, and from studies of the N2).

The complementary strengths of the theories should not, however, lead us to overlook an important difference between them, at least as they are currently framed. Specifically, Holroyd and Coles's (2002) account of the ERN differs from ours in the proposed role of the ACC: In the present model, the ERN is explained in terms of conflict monitoring by ACC, implying that this area is performing an evaluative function. In contrast, Holroyd and Coles suggest that ACC is the recipient, rather than generator, of evaluative information and that it plays a direct role in the selection of responses. That is, their theory proposes that the ERN reflects the arrival of an error-related learning signal in anterior cingulate cortex and that the role of anterior cingulate is to use this learning signal to improve performance. With regard to this issue, neuroimaging studies have provided some evidence that ACC plays an evaluative role, not an executive one (Botvinick et al., 1999; MacDonald et al., 2000). Moreover, as reviewed above, a large number of neuroimaging studies have reported ACC activity on correct trials when there is response conflict, a finding that is difficult to explain if one assumes that ACC is simply the recipient of an error signal. Thus, our working hypothesis is that ACC performs an evaluative role, monitoring for conflict during response selection. We leave open the possibility that information from conflict monitoring may be used in reinforcement learning.

The ERN as an emotional response. It has been proposed that the ERN reflects an appraisal of the emotional or motivational significance of errors, rather than reflecting the error-detection process itself (Bush et al., 2000; Gehring & Willoughby, 2002; Pailing et al., 2002). The present research has not addressed the issue of whether the ERN is a cognitive or affective correlate of errors, and our simulations do not speak directly to this question. Indeed, given that this affective processing hypothesis leaves open the issue of how errors are actually detected in the brain, it is entirely consistent with the present theory. That is, the present findings are consistent with the idea that error detection relies on conflict monitoring and that the ERN is an affective correlate of this conflict monitoring function (that is carried out in some other part of the brain). Nevertheless, as mentioned above, our working hypothesis is that ACC is responsible for error detection through conflict monitoring and that the ERN is a direct correlate of the conflict monitoring process.

Our working hypothesis is based, in part, on a consideration of the likely neural source of the ERN. It has been proposed that there are functional divisions within ACC, with dorsal-caudal regions implicated in cognitive and motor functions and more ventral-rostral regions associated with autonomic and affective function (Bush et al., 2000; Casey, Yeung, & Fosella, 2002; Devinsky et al., 1995). Previous fMRI studies have found that conflict-related activity is typically restricted to the caudal part of ACC (e.g., Botvinick et al., 1999; Braver et al., 2001; Carter et al., 1998; Kiehl et al., 2000; Menon et al., 2001), perhaps extending dorsally into the pre-SMA (Ullsperger & von Cramon, 2001). Therefore, if the ERN reflects conflict monitoring—as our theory predicts and as the co-localization of the ERN and N2 suggests—then it should be generated in caudal regions of ACC associated with cognitive and motor function.

Also relevant to the issue of whether the ERN might reflect affective consequences of errors are the findings of Nieuwenhuis et al. (2001). In a task requiring fast saccades away from a visual stimulus, participants typically make a number of errors—saccades toward the stimulus—that are quickly corrected and of which the participant remains unaware. Nieuwenhuis et al. compared ERN amplitude following these unperceived errors with ERN amplitude following errors that the participants correctly detected. This comparison is relevant to the present concerns: If the ERN reflects emotional processing of errors, then it should be absent, or at least greatly reduced, on trials in which participants remain unaware of their error. In contrast, if the ERN reflects conflict between the error and the correcting response, then an ERN should be observed on trials with unperceived errors because those errors are always corrected—a circumstance associated with high conflict. Nieuwenhuis et al.'s results were clear: ERN amplitude was as large following unperceived errors as following perceived errors, consistent with the predictions of our theory.

However, neither of the preceding lines of evidence definitively rules out the notion that the ERN is associated with affective processing. For example, regarding the localization of the ERN in caudal ACC, it might be argued that caudal ACC is involved in affective appraisals of cognitive events such as conflict or error detection (Gehring & Willoughby, 2002). Meanwhile, regarding the findings of Nieuwenhuis et al. (2001), it might be argued that participants may not be aware of their own affective responses and that the ERN to unperceived errors reflects these subconscious affective responses. Thus, although it seems plausible that participants should have greater affective responses to errors of which they are aware—particularly given how strongly they typically express their frustration in such cases—because Nieuwenhuis et al. did not directly measure affect or autonomic activity, it is impossible to rule out the hypothesis that participants in their study had strong, yet unperceived affective responses that might account for the observed ERN.

Overall, therefore, existing evidence is equivocal as to whether the ERN reflects cognitive or affective aspects of error processing. Indeed, to the extent that error detection may be inextricably linked to affective and motivational functioning, it may be impossible to separate out cognitive and emotional correlates of error detection (cf. Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Yeung, 2004). This notion is perhaps best illustrated in terms of the proposed functional role of conflict monitoring. In particular, although the present research has focused on conflict monitoring as it relates to error detection and the ERN, an important aspect of the theory concerns how information about response conflict might be used in the adaptive control of behavior (Botvinick et al., 2001). According to the theory, detection of response conflict typically leads to increased attentional focus. However, if conflict is sustained over a long period—indicating that increased effort may be insufficient to reduce conflict—participants may tend to disengage from the task (Cohen et al., 2000; Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999).

One hypothesis worthy of future investigation is that the proposed consequences of conflict detection provide an account of the functional role of affective reactions (Yeung, 2004). For example, an increase in attentional focus following conflict detection may be expressed in terms of autonomic changes related to increased alertness and arousal. Correspondingly, sustained conflict that leads to disengagement from the task may be expressed in terms of subjective feelings of frustration. On this view, it does not make sense to ask whether the ERN reflects cognitive monitoring or functions related to affect or motivation, as these are proposed to

be one and the same process. That is, conflict monitoring should not be considered to be separate from affective processing. Instead, in providing information that has direct motivational significance, conflict monitoring may provide the computational basis underlying what are observed and experienced as affective reactions.

Extensions and Future Directions

An attractive feature of the conflict monitoring theory is its computational simplicity: In our model, conflict was calculated directly as the product of the activation levels of competing response units. Much of the computational simplicity derives from the proposal that performance monitoring can rely on detecting features of processing that are reliably associated with poor performance, instead of relying on a mechanism that uses explicit information about the correct response or response accuracy. However, the feature of processing we have focused on here—the occurrence of response conflict—does rely on there being an established task set with mutually incompatible responses. The establishment of a task set is a complex process, and investigation of the mechanisms by which task sets are created, maintained, and switched is the focus of ongoing research (Allport, Styles, & Hsieh, 1994; Braver & Cohen, 2000; Gilbert & Shallice, 2002; Logan & Gordon, 2001; Rogers & Monsell, 1995; Rubinstein, Meyer, & Evans, 2001; Yeung & Monsell, 2003a, 2003b). Such issues are beyond the scope of the present research: We have been concerned solely with how one might monitor performance once a task set is established. Nonetheless, an important avenue for future research will be to determine how a task set is created such that otherwise compatible responses (e.g., key-presses with the left and right hands) are set in mutual opposition and how a conflict monitoring mechanism might be sensitive to this aspect of task set.

These challenges notwithstanding, the present research demonstrates that the conflict monitoring theory can account for observed properties of the ERN and human error-detection performance. In this way, the theory explains how findings from electrophysiological studies can be reconciled with the growing literature from neuroimaging studies showing activity in ACC associated with conditions of response competition and errors (e.g., Braver et al., 2001; Carter et al., 1998; Garavan et al., 2002; Kiehl et al., 2000; MacDonald et al., 2000; Menon et al., 2001). These studies have typically reported that caudal ACC shows the predicted sensitivity to response conflict and errors, suggesting that this region is involved in the continuous monitoring of response conflict. An intriguing possibility suggested by recent research is that other regions in the medial wall may perform more specialized conflict monitoring functions. In particular, it has been suggested that dorsal regions in SMA or pre-SMA might be selectively activated by response conflict on trials with correct responses, whereas more rostral areas in ACC might selectively respond on error trials (Braver et al., 2001; Garavan et al., 2002; Garavan, Ross, Kaufman, & Stein, 2003; Kiehl et al., 2000; Menon et al., 2001; Ullsperger & von Cramon, 2001). One possibility is that caudal ACC performs a general conflict monitoring function—and is thus activated by response conflict and errors—whereas pre-SMA and rostral ACC selectively monitor for pre- and post-response conflict, respectively, and are thus selectively activated by conflict on correct trials and by incorrect responses. This hypothesis warrants attention in future research.

The conflict monitoring theory may thus provide a unifying account of ACC activity observed in fMRI and ERP research. However, as discussed above, the theory in its present form does not attempt to explain the feedback ERN (Holroyd & Coles, 2002; Miltner et al., 1997) or the related component that is observed following late responses in experiments with response deadlines (Johnson et al., 1997; Luu et al., 2000; Pailing et al., 2000). The feedback and late-response ERNs have a similar scalp topography to the response-related ERN and, likewise, appear to be generated in the region of ACC (Miltner et al., 1997). Hence, the present theory may need to be extended in order to account for these ERP findings. We suggested above that the reinforcement learning theory proposed by Holroyd and Coles (2002) might provide a framework for understanding how information from conflict monitoring might be integrated with information from external feedback and other sources. On this account, conflict monitoring would provide just one of many inputs into the reinforcement learning process occurring in ACC, all of which generate ERN-like scalp potentials.

Any such attempt to provide a complete, unified account of ACC function may, however, need to take into account recent evidence of anatomical dissociations between regions of ACC sensitive to response errors and negative feedback (Carter, van Veen, Holroyd, Stenger, & Cohen, 2002; Gehring & Fencsik, 2001). Nevertheless, we speculate that different subregions of ACC perform related functions, all of which are responsible for evaluating internal states for evidence of breakdowns in processing and all of which can guide adjustments in control needed to improve performance. This broader view of ACC function can account for the variety of conditions under which ACC activity is observed, including response conflict and errors, negative feedback, and even pain (Craig, Reiman, Evans, & Bushnell, 1996; Peyron, Laurent, & Garcia-Larrea, 2000; Rainville, Duncan, Price, Carrier, & Bushnell, 1997; Vogt, Sikes, & Vogt, 1993). Such a class of functions would complement those responsible for monitoring the external environment for signs of threat, such as has been ascribed to the amygdala (LeDoux, 1996).

In summary, according to a broader view of ACC function, response conflict may be one valuable information source—that can provide early information about breakdowns in processing in the absence of explicit feedback—out of the many used by ACC in the evaluation of ongoing performance. The contribution of the present research is therefore to provide a computationally specified theory of one specific aspect of ACC function: monitoring of response conflict and its use in detecting errors. An important goal for future research will be to provide correspondingly detailed accounts of other proposed functions of ACC, which might include monitoring for conflict in other aspects of information processing and the processing of explicit performance feedback. One can then begin to investigate how information from response conflict monitoring might be integrated in ACC with information from these other sources and then used in the control of cognitive processing.

Conclusion

The present research has introduced a new account of the ERN, N2, and error detection in terms of the response conflict monitoring theory of anterior cingulate function. Through simulation and experiment, we have attempted to demonstrate that the conflict monitoring theory can provide a detailed account of a large corpus

of existing, and in some cases counterintuitive, findings regarding behavioral and electrophysiological phenomena related to performance monitoring. In providing an alternative to existing accounts of these phenomena, our theory attempts to answer a number of critical questions: Is the evaluative process indexed by the ERN a continuous one or is it a discrete process triggered by response execution (or some later event)? Does error detection require an explicit representation of the correct response or can it be based on detecting features of processing (such as post-response conflict) that are reliably associated with error commission? At what time might such information be available to the monitoring system, and is this timing consistent with the observed properties of the ERN? We have attempted to answer these questions by grounding our theory in a mechanistically explicit model. An important goal for future research will be to specify competing theories in comparable detail in order to provide answers to the questions raised above. Doing so will allow these accounts and ours to be compared in a rigorous, quantitative manner. However, to the extent that ours is currently the only theory that has been shown to account for the timing of the ERN, its sensitivity to the range of manipulations we simulate, and the properties of the ERP in the absence of errors, we believe that it represents a plausible account of the relationship between the ERN and error processing, one that needs to be considered alongside existing theories as we seek to understand the mechanisms underlying human performance monitoring.

References

Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 112–153). London: Routledge and Kegan Paul.

Allport, D. A. (1987). Selection for action: Some behavioural and neurophysiological considerations of attention and action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: Erlbaum.

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 421–452). Cambridge, MA: MIT Press.

Barch, D. M., Braver, T. S., Sabb, F. W., & Noll, D. C. (2000). Anterior cingulate and the monitoring of response conflict: Evidence from an fMRI study of overt verb generation. *Journal of Cognitive Neuroscience, 12*, 298–309.

Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., & Dolan, R. J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia, 9*, 907–922.

Botvinick, M. M., Braver, T. S., Carter, C. S., Barch, D. M., & Cohen, J. D. (2001). Evaluating the demand for control: Anterior cingulate cortex and crosstalk monitoring. *Psychological Review, 108*, 624–652.

Botvinick, M. M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179–181.

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex, 11*, 825–836.

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. S. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 713–737). Cambridge, MA: MIT Press.

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

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998, May 1). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*, 747–749.

Carter, C. S., van Veen, V., Holroyd, C. B., Stenger, V. A., & Cohen, J. D. (2002, June). *Errors and conflict but not error feedback engage the anterior cingulate cortex during event-related fMRI: Implications for performance monitoring in the human brain*. Paper presented at the 8th International conference on functional mapping of the human brain, Sendai, Japan.

Casey, B. J., Yeung, N., & Fosella, J. (2002). Anterior cingulate cortex. In V. S. Ramachandran (Ed.), *Encyclopedia of the human brain* (Vol. 1, pp. 145–157). Boston: Academic Press.

Cohen, J. D., Botvinick, M. M., & Carter, C. S. (2000). Anterior cingulate and prefrontal cortex: Who's in control? *Nature Neuroscience, 3*, 421–423.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex and dopamine: A connectionist approach to behaviour and biology in schizophrenia. *Psychological Review, 99*, 45–77.

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology, 105*, 239–269.

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 529–553.

Coles, M. G. H., Gratton, G., & Fabiani, M. (1990). Event-related brain potentials. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social and inferential elements* (pp. 413–455). Cambridge, UK: Cambridge University Press.

Coles, M. G. H., Scheffers, M. K., & Holroyd, C. B. (1998). Berger's dream? The error-related negativity and modern cognitive psychophysiology. In H. Witte, U. Zwiener, B. Schack, & A. Doring (Eds.), *Quantitative and topological EEG and MEG analysis* (pp. 96–102). Jena-Erlangen, Germany: Druckhaus Mayer Verlag.

Coles, M. G. H., Scheffers, M. K., & Holroyd, C. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology, 56*, 173–189.

Craig, A. D., Reiman, E. M., Evans, A. C., & Bushnell, M. C. (1996). Functional imaging of an illusion of pain. *Nature, 384*, 258–260.

Crosson, B., Sadek, J. R., Bobholz, J. A., Gokcay, D., Mohr, C. M., Leonard, C. M., et al. (1999). Activity in the paracingulate and cingulate sulci during word generation: An fMRI study of functional anatomy. *Cerebral Cortex, 9*, 307–316.

Davies, P. L., Segalowitz, S. J., Dywan, J., & Pailing, P. E. (2001). Error-negativity and positivity as they relate to other ERP indices of attentional control and stimulus processing. *Biological Psychology, 56*, 191–206.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science, 5*, 303–305.

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

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of target letters in a non-search task. *Perception and Psychophysics, 16*, 143–149.

Eriksen, C. W., Coles, M. G. H., Morris, L. R., & O'Hara, W. P. (1985). An electromyographic examination of response competition. *Bulletin of the Psychonomic Society, 23*, 165–168.

Falkenstein, M., Hohnsbein, J., & Hoorman, J. (1995). Event-related potential correlates of errors in reaction tasks. In G. Karmos, M. Molnar,

V. Csepe, I. Czigler, & J. E. Desmedt (Eds.), *Perspectives of event-related potentials research* (pp. 287–296). Amsterdam: Elsevier.

Falkenstein, M., Hohnsbein, J., Hoorman, J., & Blanke, L. (1990). Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In C. H. M. Brunia, A. W. K. Gaillard, & A. Kok (Eds.), *Psychophysiological brain research* (Vol. 1, pp. 192–195). Tilburg, the Netherlands: Tilburg University Press.

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

Falkenstein, M., Hoorman, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology, 51,* 87–107.

Garavan, H., Ross, H., Kaufman, T. J., & Stein, E. A. (2003). A midline dissociation between error processing and response conflict monitoring. *NeuroImage, 20,* 1132–1139.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: Inhibition, error detection, and correction. *Neuroimage, 17,* 1820–1829.

Gehring, W. J., & Fencsik, D. (1999, April). *Slamming on the brakes: An electrophysiological study of error response inhibition.* Paper presented at the annual meeting of the Cognitive Neuroscience Society, Washington, DC.

Gehring, W. J., & Fencsik, D. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *Journal of Neuroscience, 21,* 9430–9437.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4,* 385–390.

Gehring, W. J., Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Probability effect on stimulus evaluation and response processes. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 198–216.

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive–compulsive disorder. *Psychological Science, 11,* 1–6.

Gehring, W. J., & Willoughby, A. R. (2002, March 22). The medial frontal cortex and the rapid processing of utility information. *Science, 295,* 2279–2282.

Gemba, H., Sasaki, K., & Brooks, V. B. (1986). "Error" potentials in limbic cortex (anterior cingulate area 24) of monkeys during motor learning. *Neuroscience Letters, 70,* 223–227.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology, 44,* 297–337.

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121,* 480–506.

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 331–344.

Greenblatt, R. E., & Robinson, S. E. (1994). A simple head shape approximation for the 3 shell model. *Brain Topography, 6,* 331.

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Henninghausen, E. (2000). N200 in the Eriksen-Task: Inhibitory executive processes? *Journal of Psychophysiology, 14,* 218–225.

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., Dien, J., & Coles, M. G. H. (1998). Error-related scalp potentials elicited by hand and foot movements: Evidence for an output-independent error-processing system in humans. *Neuroscience Letters, 242,* 65–68.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *NeuroReport, 14,* 2481–2484.

Holroyd, C. B., Praamstra, P., Plat, E., & Coles, M. G. H. (2002). Spared error-related potentials in mild to moderate Parkinson's disease. *Neuropsychologia, 40,* 2116–2124.

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2004). *A mechanism for error detection in speeded response time tasks*. Manuscript submitted for publication.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA, 79,* 2554–2558.

Johnson, T. M., Otten, L. J., Boeck, K., & Coles, M. G. H. (1997). Am I too late? The neural consequences of missing a deadline. *Psychophysiology, 34,* S48.

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

Kok, A. (1986). Effects of degradation of visual stimuli on components of the event-related potential (ERP) in go/nogo reaction tasks. *Biological Psychology, 23,* 21–38.

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology, 33,* 282–294.

Lange, J. J., Wijers, A. A., Mulder, L. J. M., & Mulder, G. (1998). Color selection and location selection in ERPs: Differences, similarities and "neural specificity". *Biological Psychology, 48,* 153–182.

LeDoux, J. E. (1996). *The emotional brain*. New York: Simon & Schuster.

Leuthold, H., & Sommer, W. (1999). ERP correlates of error processing in spatial S-R compatibility tasks. *Clinical Neurophysiology, 110,* 342–357.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color–word interference effect. *Neuropsychologia, 38,* 701–711.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108,* 393–434.

Luu, P., Flaisch, T., & Tucker, D. M. (2000). Medial frontal cortex in action monitoring. *Journal of Neuroscience, 20,* 464–469.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000, June 9). Dissociating the role of dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288,* 1835–1838.

MacKay, D. G. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition, 1,* 199–225.

Maylor, E. A., & Rabbitt, P. M. A. (1987). Effects of alcohol and practice on choice reaction time. *Perception and Psychophysics, 42,* 465–475.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62,* 203–208.

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping, 12,* 131–143.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related 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.

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: Erlbaum.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to aware-

ness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38,* 752–760.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a Go/NoGo task: Effects of response conflict and trial-type frequency. *Cognitive, Affective, and Behavioral Neuroscience, 3,* 17–26.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). New York: Plenum.

Pailing, P. E., Segalowitz, S. J., & Davies, P. L. (2000). Speed of responding and the likelihood of error-like activity in correct trial ERPs. *Psychophysiology, 37,* S76.

Pailing, P. E., Segalowitz, S. J., Dywan, J., & Davies, P. L. (2002). Error negativity and response control. *Psychophysiology, 39,* 198–206.

Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences of the USA, 87,* 256–259.

Paus, T., Koski, L., Caramanos, Z., & Westbury, C. (1998). Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies. *NeuroReport, 9,* R37–R47.

Paus, T., Petrides, M., Evans, A. C., & Meyer, E. (1993). Role of human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of Neurophysiology, 20,* 453–469.

Peyron, R., Laurent, B., & Garcia-Larrea, L. (2000). Functional imaging of brain responses to pain. A review and meta-analysis. *Neurophysiology Clinique–Clinical Neurophysiology, 30,* 263–288.

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology, 60,* 423–434.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77,* 97–131.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1992). *Numerical recipes in C* (2nd Edition). Cambridge, UK: Cambridge University Press.

Pritchard, W. S., Shappell, S. A., & Brandt, M. E. (1991). Psychophysiology of N200/N400: A review and classification scheme. In J. R. Jennings, P. K. Ackles, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 4, pp. 43–106). London: Jessica Kingsley.

Rabbitt, P. M. A. (1966). Error correction time without external error signals. *Nature, 212,* 438.

Rabbitt, P. M. A. (1967). Time to detect errors as a function of factors affecting choice-reaction time. *Acta Psychologica, 27,* 131–142.

Rabbitt, P. M. A. (1968). Three kinds of error-signalling responses in a serial choice task. *Quarterly Journal of Experimental Psychology, 20,* 179–188.

Rabbitt, P. M. A. (1990). Age, IQ and awareness, and recall of errors. *Ergonomics, 33,* 1291–1305.

Rabbitt, P. M. A. (2002). Consciousness is slower than you think. *Quarterly Journal of Experimental Psychology, 55A,* 1081–1092.

Rabbitt, P. M. A., Cumming, G., & Vyas, S. M. (1978). Some errors of perceptual analysis in visual search can be detected and corrected. *Quarterly Journal of Experimental Psychology, 30,* 319–332.

Rabbitt, P. M. A., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29,* 727–743.

Rabbitt, P. M. A., & Vyas, S. M. (1981). Processing a display even after you make a response to it. How perceptual errors can be corrected. *Quarterly Journal of Experimental Psychology, 33A,* 223–239.

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997, August 15). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science, 277,* 968–971.

Ritter, W., Simson, R., Vaughan, H. G., & Friedman, D. (1979, March 30). A brain event related to the making of a sensory discrimination. *Science, 203,* 1358–1361.

Ritter, W., Simson, R., Vaughan, H. G., & Macht, M. (1982, November 26). Manipulation of event-related potential manifestations of information processing stages. *Science, 218,* 909–911.

Rodríguez-Fornells, A., Kurzbuch, A. R., & Münte, T. F. (2002). Time course of error detection and correction in humans: Neurophysiological evidence. *Journal of Neuroscience, 22,* 9990–9996.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., et al. (2001). Mapping motor inhibition: Conjunctive brain activations across different versions of go/no-go and stop tasks. *NeuroImage, 13,* 250–261.

Rubinstein, J., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 763–797.

Rush, S., & Driscoll, D. A. (1968). Current distributions in the brain from surface electrodes. *Anesthesia and Analgesia, 47,* 717–723.

Scheffers, M. K. (1999). *Performance monitoring: Error detection and the error-related negativity in choice–reaction time tasks*. Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign.

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgements of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 141–151.

Scheffers, M. K., Coles, M. G. H., Bernstein, P., Gehring, W. J., & Donchin, E. (1996). Event-related potentials and error-related processing: An analysis of incorrect responses to go and no-go stimuli. *Psychophysiology, 33,* 42–53.

Schultz, W., Dayan, P., & Montague, P. R. (1997, March 14). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Servan-Schreiber, D., Bruno, R. M., Carter, C. S., & Cohen, J. D. (1998). Dopamine and the mechanisms of cognition: Part I. A neural network model predicting dopamine effects on selective attention. *Biological Psychiatry, 43,* 713–722.

Servan-Schreiber, D., Carter, C. S., Bruno, R. M., & Cohen, J. D. (1998). Dopamine and the mechanisms of cognition: Part II. D-amphetamine effects in human subjects performing a selective attention task. *Biological Psychiatry, 43,* 723–729.

Spencer, K. M., & Coles, M. G. H. (1999). The lateralized readiness potential: Relationship between human data and response activation in a connectionist model. *Psychophysiology, 36,* 364–370.

Stuphorn, V., Taylor, T. L., & Schall, J. D. (2000). Performance monitoring by the supplementary eye field. *Nature, 408,* 857–860.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Taylor, S. F., Kornblum, S., Minoshima, S., Oliver, L. M., & Koeppe, R. A. (1994). Changes in medial cortical blood flow with a stimulus-response compatibility task. *Neuropsychologia, 32,* 249–255.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the USA, 94,* 14792–14797.

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.

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999, January 22). The role of locus coeruleus in the regulation of cognitive performance. *Science, 283,* 549–554.

van Veen, V., & Carter, C. S. (2002). The timing of action monitoring in rostral and caudal anterior cingulate cortex. *Journal of Cognitive Neuroscience, 14,* 593–602.

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the "error negativity" specific to errors? *Biological Psychology, 51,* 109–128.

Vogt, B. A., Sikes, R. W., & Vogt, L. J. (1993). Anterior cingulate cortex and the medial pain system. In B. A. Vogt & M. Gabriel (Eds.), *Neurobiology of cingulate cortex and limbic thalamus: A comprehensive handbook* (pp. 313–344). Boston: Birkhauser.

Yeung, N. (2004). Relating cognitive and affective theories of the error-related negativity. In M. Ullsperger & M. Falkenstein (Eds.), *Errors, conflicts, and the brain. Current opinions on performance monitoring* (pp. 63–70). Leipzig, Germany: Max Planck Institute of Cognitive Neuroscience.

Yeung, N., Holroyd, C. B., & Cohen, J. D. (in press). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex.*

Yeung, N., & Monsell, S. (2003a). The effects of recent practice on task switching. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 919–936.

Yeung, N., & Monsell, S. (2003b). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 255–269.

Appendix

Simulation Details

On each trial, the network was run for 3 preparatory cycles, then for 50 further cycles. On each cycle, the activation of each unit was calculated according to its net input. The net input to each unit of the behavioral network was given by:

$$\text{net}_i = (\text{ext}_i * \text{estr}) + \sum \text{act}_j w_{ij} \text{s} + \text{noise},$$

where ext_i is the external input to the unit, estr is a constant scaling the external input to each unit (set equal to 0.4), act_j is the activation of the sending unit on the previous time step, w_{ij} is the weight of the connection between the two units, and s is a scaling parameter, set to 0.08 for excitatory weights and 0.12 for inhibitory weights. Noise is normally distributed with a mean of 0.00 and a standard deviation of 0.035.

Net inputs were initialized to zero at the start of each trial. An external input of 0.03 was supplied to the response units during 3 preparatory cycles. The network was then given external inputs corresponding to the stimulus and attentional input (as well as maintaining $\text{ext}_i = 0.03$ for response units). ext_i was set to 0.15 for all relevant stimulus units. The input to the three attentional units was modulated across trials according to the degree of conflict experienced on previous trials. Specifically, ext_i to the center attentional unit was given by:

$$\text{ext}_{\text{center}} = \lambda \text{ext}_{\text{center}}(t-1) + (1-\lambda)[\alpha \text{E}(t-1) + \beta],$$

where $\text{ext}_{\text{center}}(t-1)$ is the external input to the center attention unit on the previous trial, λ, α, and β are constants set to 0.5, 4.41, and 1.08, respectively, and $\text{E}(t-1)$ is the total energy (Hopfield, 1982) in the response layer on the previous trial. The energy (conflict) at each time step, t, was calculated as $-\sum\sum \text{act}_i * \text{act}_j * w_{ij}$, where i and j are indexed over all units in the response layer, giving:

$$\text{energy}_t = -2 * (\text{act}_{r,H,t} * \text{act}_{r,S,t} * -3),$$

where $\text{act}_{r,H,t}$ and $\text{act}_{r,S,t}$ are the activation levels of the response units corresponding to the H and S stimuli, respectively, at time step t. Thus, energy is calculated as the product of activation of the two response units at time step t, multiplied by the strength of the lateral inhibition between them. Energy, constrained to be ≥ 0, was summed across all time steps of a trial to give $\text{E}(t)$.

$\text{ext}_{\text{center}}$ was constrained to lie between 1 and 3. ext_i to flanker attention units was given by:

$$\text{ext}_{\text{flanker}} = (3 - \text{ext}_{\text{center}})/2.$$

The activation of a unit was calculated from its net input as follows:

If $\text{net}_i > 0$, the change in activation on that time step was given by:

$$\Delta\text{act}_i = [(\text{act}_{\max} - \text{act}_i) * \text{net}_i] - [(\text{act}_i - \text{act}_{\text{rest}}) * \text{decay}].$$

If $\text{net}_i < 0$, the change in activation was given by:

$$\Delta\text{act}_i = [(\text{act}_i - \text{act}_{\min}) * \text{net}_i] - [(\text{act}_i - \text{act}_{\text{rest}}) * \text{decay}],$$

where $\text{act}_{\max}$, $\text{act}_{\min}$, and act_{rest} are the maximum, minimum, and resting activations of the units, set to 1.0, −0.2, and −0.1, respectively. Decay was a constant set to 0.1. If $\text{act}_i > \text{act}_{\max}$, act_i was set equal to $\text{act}_{\max}$. Similarly, if $\text{act}_i < \text{act}_{\min}$, act_i was set equal to $\text{act}_{\min}$. The gain of a unit was manipulated by multiplying its net input by a constant scaling factor. Multiplying the net input by a constant value greater than 1 increases the rate at which the unit approaches $\text{act}_{\max}$ (or $\text{act}_{\min}$, if the net input to the unit is inhibitory), thus capturing the notion that gain of the unit has increased. Scaling the net input by a value less than 1 results in a corresponding reduction in the rate at which $\text{act}_{\max}$ is approached, resulting in reduced gain.

A response was recorded if the activation of either response unit exceeded a prespecified response threshold (0.18, except where noted). The model continued to process until the end of the 50 cycle run, regardless of the time at which the response was made. However, external input to the model was stopped after a smaller number of cycles ($M = 6.0$ cycles; $SD = 0.5$ cycles). This was done to keep the amount of post-response processing relatively constant across trials with different RTs and was also used to simulate the idea that participants in experiments do not continue to process indefinitely after they have responded.

There were excitatory connections between layers and inhibitory connections within layers of the network. The connection weights, except where noted in the text, were as follows: feedforward excitatory connections from stimulus to response units = 1.5; bidirectional excitatory connections between the stimulus and attention units = 2.0; stimulus layer lateral inhibition = −2.0; response layer lateral inhibition = −3.0; attention layer lateral inhibition = −1.0. Note that each stimulus unit had mutual inhibitory connections with all other stimulus units. Similarly, each attentional unit laterally inhibited both other attentional units.

Received August 17, 2001
Revision received August 21, 2003
Accepted October 29, 2003 ■