Behavioral/Systems/Cognitive

Decision Processes in Human Performance Monitoring

Marco Steinhauser[1] and Nick Yeung[2]

[1]Department of Psychology, University of Konstanz, D-78457 Konstanz, Germany, and [2]Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom

The ability to detect and compensate for errors is crucial in producing effective, goal-directed behavior. Human error processing is reflected in two event-related brain potential components, the error-related negativity (Ne/ERN) and error positivity (Pe), but the functional significance of both components remains unclear. Our approach was to consider error detection as a decision process involving an evaluation of available evidence that an error has occurred against an internal criterion. This framework distinguishes two fundamental stages of error detection—accumulating evidence (input), and reaching a decision (output)—that should be differentially affected by changes in internal criterion. Predictions from this model were tested in a brightness discrimination task that required human participants to signal their errors, with incentives varied to encourage participants to adopt a high or low criterion for signaling their errors. Whereas the Ne/ERN was unaffected by this manipulation, the Pe varied consistently with criterion: A higher criterion was associated with larger Pe amplitude for signaled errors, suggesting that the Pe reflects the strength of accumulated evidence. Across participants, Pe amplitude was predictive of changes in behavioral criterion as estimated through signal detection theory analysis. Within participants, Pe amplitude could be estimated robustly with multivariate machine learning techniques and used to predict error signaling behavior both at the level of error signaling frequencies and at the level of individual signaling responses. These results suggest that the Pe, rather than the Ne/ERN, is closely related to error detection, and specifically reflects the accumulated evidence that an error has been committed.

Introduction

The detection of errors through continuous monitoring of action outcomes is crucial for achieving optimal performance. Error monitoring is associated with activity in a widespread network of brain regions, including areas in the medial and lateral prefrontal and parietal cortices (Kiehl et al., 2000; Ullsperger and von Cramon, 2001; Ridderinkhof et al., 2004). Scalp EEG methods provide important evidence about this neural activity because errors in simple choice tasks are known to elicit an early negative deflection called the error negativity (Ne) (Falkenstein et al., 1990) or error-related negativity (ERN) (Gehring et al., 1993) that is followed by a later positive deflection called the error positivity (Pe) (Falkenstein et al., 1990).

The Ne/ERN and Pe have received substantial scrutiny, yet fundamental questions remain. In particular, while it is often assumed that the two components reflect sequential stages of error processing (Falkenstein et al., 1990; van Veen and Carter, 2002), there is little consensus about the precise nature of these stages. Either or both components could reflect precursors to explicit error detection (like conflict monitoring, Yeung et al., 2004), the error detection process itself (Nieuwenhuis et al., 2001), evaluation of the significance of a detected error (Hajcak et al., 2005), or initiation of subsequent behavioral adjustments

(Holroyd and Coles, 2002). The present research aimed to provide new insight into this issue, using a novel approach to address a simple, but critical, unanswered question: Do the Ne/ERN and Pe reflect neural signals indicating that an error has occurred, or do they rather reflect earlier processes that might provide the input to such an error detection system?

Our approach was to treat error monitoring as a decision process in which participants make judgments ("Did I respond correctly?") on the basis of imperfect evidence. Signal detection theory (Green and Swets, 1966) and evidence accumulation models (Ratcliff and Rouder, 1998) provide well developed formalisms for understanding such decisions. We leveraged key ideas from these approaches—in particular, the distinction between accumulated decision evidence and categorical decision output—to investigate the informational content of error-related EEG components. Experimental participants were asked to signal the errors they made while performing a difficult perceptual discrimination task. By varying the incentives associated with accurate error signaling, we manipulated participants' signaling criterion. As detailed below, neural signals related to decision evidence and decision output should vary in markedly different ways as a function of the criterion applied.

Looking ahead, our results suggest that even the later Pe component is better characterized as reflecting accumulated decision evidence—the input to decisions about response accuracy—rather than decision output. To evaluate this idea further, we used multivariate machine learning techniques to establish robust measures of error-related activity on individual trials (Parra et al., 2002; Philiastides and Sajda, 2006, 2007; Philiastides et al., 2006). Using these techniques, we investigated the degree to

which participants' observed error decisions could be predicted based on neural error signals preceding these decisions.

Materials and Methods

Participants

Sixteen right-handed participants (14 female) between 18 and 23 years of age (mean 19.4) with normal or corrected-to-normal vision participated in the study. Participants were recruited from the Oxford University community for course credit or payment, and were paid an additional performance-dependent bonus.

Task and procedure

The present study aimed to manipulate participants' error signaling decisions (Rabbitt, 1968; Steinhauser et al., 2008) through the use of performance incentives, and to investigate the impact of this manipulation on error-related EEG components. Participants performed the same "primary" task throughout the experiment—a difficult perceptual discrimination task—but across conditions were differentially rewarded according to the accuracy with which they signaled their errors in this primary task. In one condition, participants were punished most strongly (i.e., lost most money) when they failed to signal errors they had made in the discrimination task; in the other, they were punished most strongly when they signaled an error after having actually made a correct response. These incentives were designed to modulate participants' internal criterion for signaling errors: In the first condition, they should adopt a low criterion to avoid "misses;" in the second, they should adopt a high criterion to avoid "false alarms." Critically, varying the criterion in this way should have different effects on decision evidence and decision output: If participants adopt a low criterion, they should frequently signal errors, but many of those errors will be signaled on the basis of weak evidence that an error has occurred. In contrast, adopting a high criterion should reduce the number of error signals, but these signaled errors should be associated with stronger evidence. As detailed below, this leads to very different predictions as to how the criterion should influence error-related EEG components depending on whether they reflect decision evidence or decision output.

On each trial, participants first performed a difficult perceptual discrimination and then were prompted to make a signaling response when they thought they had made an error. The discrimination task required participants to decide which of two boxes presented on a screen was brighter. The boxes were noisy and the brightness difference rather small. In this way, stimulus uncertainty was induced, which is known to impair error processing (Pailing and Segalowitz, 2004) and to produce undetected errors (Scheffers and Coles, 2000). This design allowed us not only to control the absolute number of errors but also the rate of undetected errors, which is an important precondition to directly manipulate error detection.

A sample trial is depicted in Figure 1. First, a white fixation cross was centrally presented for 500 ms. A stimulus then appeared for 160 ms, followed by a blank screen. The stimulus consisted of two boxes presented on a black background above and below a white fixation cross. The boxes each consisted of a 64-by-64 array of randomly arranged white and black pixels, with new arrays generated on each trial. Discrimination difficulty depended on the relative proportions of white and black pixels in the two boxes. The difficulty level was set individually for each participant (for details see supplemental materials, available at www.jneurosci.org). Participants responded by pressing one of two keys on a standard keyboard: the "T" key with the left index finger when the upper box was brighter and the "G" key with the right index finger when the lower box was brighter. Five hundred milliseconds after the response, the word "error?" was centrally presented for 1000 ms. During that time, participants were instructed to press the space bar with their right thumb if they thought that they had committed an error in the primary task. Another blank
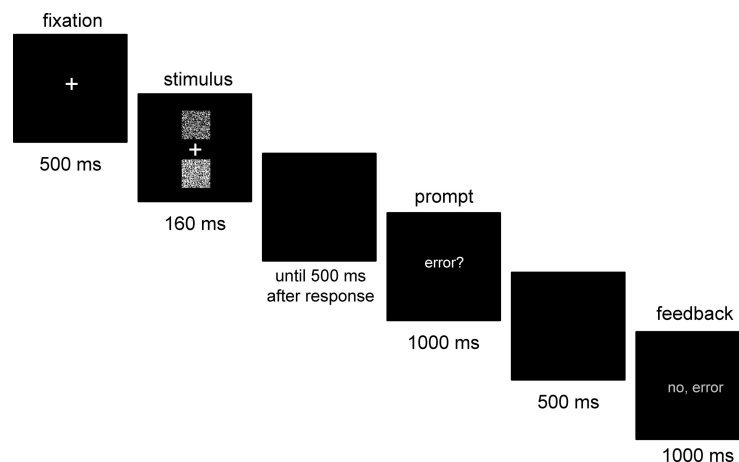

Figure 1. Sequence of stimulus events in a typical trial. Participants were first required to indicate which of two boxes in the stimulus was brighter. Following the error prompt, they pressed a signaling key if they judged that their primary task response was incorrect.

screen then appeared for 500 ms, followed by a feedback screen presented for 1000 ms.

The feedback screen indicated the accuracy of both their primary task response and their error signaling response. If the primary task response was correct, and was not followed by an error signaling response, the feedback indicated "yes, correct" in green (correct rejection). If the primary task response was correct, but was followed by an erroneous error signal, the feedback indicated "no, correct" in red (false alarm). If an incorrect primary task response was followed by an error signaling response, the feedback indicated "yes, error" in green (hit). Finally, if an incorrect primary task response remained unsignaled, the feedback indicated "no, error" in red (miss). In experimental blocks, the feedback screen additionally indicated the amount of win or loss (e.g., "+2p" or "−2p"), determined according to the incentive scheme described below.

The experiment consisted of two sessions on consecutive days. In the practice session on the first day, the task was introduced and task difficulty was calibrated using a staircase method to obtain a reasonable number of signaled and unsignaled errors (see supplemental materials, available at www.jneurosci.org, for details). The test session began with a brief reminder practice, followed by experimental blocks in which we introduced the critical manipulation of variable incentives for error signaling. Specifically, before the first experimental block, participants were told that they would receive a reward based on the accuracy of their error signaling responses. Two reward schemes were chosen so as to manipulate the participants' decision criterion for error signaling. In the high criterion condition, participants lost 2 points for each miss, but lost 10 points for each false alarm, an incentive scheme designed to bias participants toward signaling only if they were highly confident that they had made an error. In the low criterion condition, in contrast, participants lost 10 points for each miss and only 2 points for each false alarm, an incentive scheme that should encourage participants to signal more errors. In both conditions, 2 points were earned for correct rejections and hits. At the end of the experiment, points were converted into a monetary reward (1 point = 1 pence).

When a new condition was introduced, the reward scheme was presented on the screen and participants were additionally instructed that, to maximize their reward, they should "avoid pressing the signaling key in case of a correct response" (high criterion) or "avoid not pressing the signaling key in case of an error" (low criterion). In addition, participants were reminded before each block that they should maintain their response speed. This was done to maintain similar levels of primary task performance across conditions. In particular, we anticipated that participants might respond more carefully in the condition in which errors (specifically, unsignaled errors) tended to be punished—the low criterion condition. As described below, the instructions appear to have been successful: Participants responded if anything a little more quickly in this condition.

Each condition comprised six consecutive blocks, two practice blocks of 30 trials each followed by four experimental blocks of 60 trials. Thus,

240 trials per condition were included in the data analysis. The order of the two conditions was counterbalanced across participants.

EEG data acquisition

During the experiment, participants were seated in a dimly lit, electrically shielded room. The electroencephalogram (EEG) was recorded during the test session using Ag-AgCl electrodes from channels FP1, FPz, FP2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, POz, O1, Oz, O_2 as well as the right mastoid. In addition, the vertical and horizontal electrooculogram (EOG) was recorded from electrodes above and below the right eye and on the outer canthi of both eyes. All electrodes were referenced to the left mastoid and off-line re-referenced to linked mastoids. Electrode impedances were kept below 5 kΩ. EEG and EOG data were continuously recorded using SynAmps2 amplifiers (Neuroscan) at a sampling rate of 1000 Hz, with a gain of 2816 and 29.8 nV resolution.

Data analysis

Behavioral data. We first analyzed the behavioral data to test whether the manipulation of signaling criterion was successful. On the basis of the accuracy of the initial response and the occurrence of a signaling response, trials were categorized as correct rejections (correct initial response, no signaling response), false alarms (correct initial response, signaling response), misses (wrong initial response, no signaling response), and hits (wrong initial response, signaling response). The absolute frequencies were used to calculate the hit rate (H = proportion of hits among all errors) and the false alarm rate (FA = proportion of false alarms among all correct trials) for both conditions. We then estimated two parameters from signal detection theory (Green and Swets, 1966; Macmillan and Creelman, 1991): the detection criterion, c, and the sensitivity, d', which provides a criterion-independent measure of detection performance.

Event-related potential analyses. Predictions from a simple decision making model were tested by analyzing event-related potential (ERP) components, specifically focusing on the Ne/ERN and Pe occurring time-locked to responses in the primary task (i.e., shortly before participants made their overt error signaling response). EEG data preprocessing began with the correction of ocular artifacts using a regression approach (Semlitsch et al., 1986). Epochs were then extracted for a period from 500 ms before until 1000 ms after each primary task response. Baseline activity was removed by subtracting the average voltage in an interval from 150 to 50 ms before the response. Trials with voltages of 50 μV below or above the mean were excluded. Pe amplitude was quantified as the difference between error and correct trial waveforms in an interval from 250 to 350 ms after the response. The Ne/ERN was quantified as the difference between error and correct trial waveforms from -10 to 90 ms relative to the response. These difference values were computed for each channel. However, based on topographical information, statistical analysis was applied only to data from channel CPz for the Pe and from channel FCz for the Ne/ERN.

Single-trial analysis. As described below, manipulation of error signaling incentive primarily affected the Pe component and had little effect on the Ne/ERN. Our follow-up analyses therefore focused on the Pe. In particular, we attempted to estimate the amplitude of the Pe on individual trials and then correlate the amplitude of this component with participants' error signaling decisions. To achieve this, we used the linear integration method introduced by Parra et al. (2002) to measure error-related EEG activity with improved signal-to-noise ratio. The rationale of this method is to extract a specific spatial component of the ERP by constructing a classifier that maximally discriminates between two conditions differing in this component. Specifically, with $x(t)$ being the vector of electrode activity at time t, we used logistic regression to compute a spatial weighting coefficient v such that the component,

$$y(t) = v^T x(t),$$

is maximally discriminating between two different conditions, occurring at different times t. Thus, the improvement in single-trial signal-to-noise ratio is achieved by combining data across electrodes (rather than across trials as in conventional ERP analyses). In the present case, we used this method to discriminate between error and correct-response trials to estimate error-related EEG activity on individual trials (independent of criterion condition). As input, we used T samples from each of the N baseline-corrected ERP epochs, resulting in a training set of size NT. After finding the optimal v, we estimated the error signal, $\bar{y}_k$, on each trial k by averaging across the T samples from each trial. This value ranges between 0 and 1, with higher values indicating a higher probability that the trial was an error.

To visualize the spatial distribution of weights of the discriminating component, we computed the coupling coefficient vector,

$$a = \frac{Xy}{y^T y},$$

with time t being a dimension of the matrix X and the vector y. Coupling coefficients represent the activity at each electrode site that correlates with the discriminating component, and thus can be thought of as the "sensor projection" of that component (Parra et al., 2002, 2005).

To ensure that the error signal extracted with this method is equivalent to the Pe, we analyzed the time course and sensor projection of the discriminating component. To this end, the analysis was applied in a moving window (width = 100 ms) along a range from -400 to 600 ms relative to the response. Classifier sensitivity was quantified in terms of Az score, which corresponds to the area under the Receiver Operating Characteristic curve, and for which 0.5 indicates chance-level classification and 1 indicates perfect discrimination. Az scores were computed for each window using split-half cross-validation, i.e., the classifier was trained on half of the trials and was then used to predict the category (correct or error) on the remaining trials. This procedure was repeated for each half of 10 random splits, and the average of these 20 values was taken as the overall sensitivity for a specific window and participant. Variation in sensitivity across time points was used as an estimator of the time course of the component (Parra et al., 2005). To test whether sensitivity at each time point significantly exceeded chance level, a permutation test was applied (Philiastides et al., 2010). For each time point and participant, a test distribution under the null hypothesis was generated by recomputing Az scores with random assignment of truth labels (i.e., random assignment of each trial to the correct/error categories). This procedure was repeated 100 times for each of the 20 subsets of trials from which each Az score was computed. The resulting 2000 values represented the test distribution, and were used to determine critical Az values associated with significance levels of 0.05 and 0.01. Overall critical Az values were computed by averaging across participants.

Results

Behavioral data

Error signaling performance

Behavioral data were analyzed using techniques from signal detection theory (Green and Swets, 1966; Macmillan and Creelman, 1991) to test whether we were successful in manipulating participants' error detection criterion. The results are presented in Table 1. We first calculated the hit rate and false alarm rate for the two conditions. Higher rates of both hits and false alarms were observed in the low criterion condition compared with the high criterion condition, $t_{(15)} = 7.26$ and 9.27, $p < 0.001$, indicating that participants made more error signaling responses in the former. To examine whether these differences indeed reflected a shift in signaling criterion, we calculated estimates of sensitivity (d') and detection criterion (c). As expected, criterion was larger in the high criterion condition than in the low criterion condition, $t_{(15)} = 9.99$, $p < 0.001$, indicating a bias toward more frequent signaling in the latter. In addition, we found a nonsignificant trend toward an increased sensitivity in the low criterion condition compared with the high criterion condition, $t_{(15)} = 1.95$, $p < 0.10$, indicating that participants were slightly better at discriminating between errors and corrects in the low criterion condition. However, because sensitivity is proportional

Table 1. Behavioral performance: error signaling rates, estimated signal detection parameters, and primary task error rates and response times for the two criterion conditions

	Low Criterion		High Criterion	
	Mean	SEM	Mean	SEM
Error signaling performance				
Hit rate (%)	82.1	0.03	43.3	0.04
False alarm rate (%)	21.0	0.02	4.6	0.00
Criterion c	−0.07	0.07	1.02	0.07
Sensitivity d'	1.82	0.11	1.66	0.13
Primary task performance				
Error rate (%)	35.0	0.02	31.9	0.01
RT false alarms (ms)	435	27	436	29
RT Corr. Rej. (ms)	428	23	425	20
RT hits (ms)	381	31	372	21
RT misses (ms)	440	36	432	27

RT, Response time; Corr. Rej., correct rejections. Hit rate is the relative frequency of hits relative to all errors. False alarm rate is the relative frequency of false alarms relative to all correct trials.

to the difference between hit rate and false alarm rate, this trend might simply reflect a floor effect in false alarm rates for the high criterion condition. Together, these results clearly show that our manipulation of the detection criterion was successful.

Primary task performance
We additionally analyzed error rates and response times (RTs) in the primary brightness discrimination task. Error rates were somewhat greater in the low criterion condition than in the high criterion condition, although this difference failed to reach significance, $t_{(15)} = 1.94$, $p < 0.10$. This effect is somewhat surprising because, on average, errors were associated with a higher monetary loss in the low criterion condition than in the high criterion condition (because in the latter it was false alarms that were primarily punished). However, the result demonstrates that our efforts were successful in counteracting any bias toward participants producing fewer errors in the low criterion condition. Mean RTs for each trial type were entered into a three-way repeated-measures ANOVA with factors of condition (high criterion, low criterion), accuracy (correct, error) and signaling (signaled, unsignaled). This analysis revealed no reliable effects of incentive condition (all $Fs < 1$), indicating that primary task performance was little affected by our manipulation of error signaling incentives. A significant interaction between accuracy and signaling, $F_{(1,15)} = 17.2$, $p < 0.001$, indicated that whereas similar RTs occurred for misses (436 ms), false alarms (435 ms) and correct rejections (426 ms), primary task responses were faster for hits (i.e., signaled errors, 376 ms). This finding supports the assumption that detected errors mostly occur when participants respond prematurely before they have fully processed the stimulus (Scheffers and Coles, 2000).

ERP analysis: Ne/ERN and Pe
The central aim of the present study was to identify the stages of error processing reflected in the Ne/ERN and Pe. We specifically focused on the question of whether these error-related EEG components emerge before or after decisions about response accuracy are reached: That is, do either or both components reflect the accumulation of available evidence that an error has occurred, or do they reflect the output of this decision process? Although this theoretical distinction appears straightforward, in practice the component stages are difficult to dissociate empirically because decision outputs are strongly dependent on the strength of decision evidence. For example, both evidence strength and the frequency of error decisions should be reduced if the difficulty of the primary task is increased (e.g., by increasing stimulus uncertainty) (Pailing and Segalowitz, 2004) because this manipulation should impair participants' ability to derive a representation of the correct response. Similarly, neural correlates of both evidence strength and decision output should be increased on trials on which errors are detected than on trials with undetected errors.

In previous research, only the Pe has consistently been found to increase for detected errors (Nieuwenhuis et al., 2001; Endrass et al., 2005, 2007; O'Connell et al., 2007; Shalgi et al., 2009; for an overview, see Ullsperger et al., 2010). For the Ne/ERN, this relationship seems to depend on how undetected errors are induced and whether detected and undetected errors differ with respect to certain task-related features (such as post-error response conflict resulting from corrective response activation) (Yeung et al., 2004). A decreased Ne/ERN for undetected errors has been observed in tasks in which undetected errors are caused by data limitations induced by stimulus degradation (Scheffers and Coles, 2000) or masking (Maier et al., 2008). In contrast, no such decrease has been found in antisaccade tasks, in which even undetected errors may be associated with strong corrective response activation, as indicated by high rates of overt correction following these errors (Nieuwenhuis et al., 2001; Endrass et al., 2007). Since the present task is more similar to the former cases, it is perhaps not surprising that we obtained an increase in both the Ne/ERN, $t_{(15)} = 2.18$, $p < 0.05$, and Pe, $t_{(15)} = 21.2$, $p < 0.001$, for hits compared with misses in the high criterion condition (in which a sufficient number of misses occurred to allow a robust comparison). However, although this result suggests a clear relationship between the Ne/ERN, Pe, and error detection in this task, it does not identify the precise stage of error detection reflected in these components. Thus, the present study aimed to manipulate the error detection process more directly to determine whether the components reflect early or later stages of error processing.

As described above, our critical manipulation was of participants' error detection criterion. Figure 2 presents a schematic illustration of our experimental logic. Following standard models of decision making (Green and Swets, 1966; Ratcliff and Rouder, 1998), we assume that on each trial the monitoring system accumulates evidence for an error which, due to uncertainty, varies across trials (Fig. 2A, left). An error is detected if the evidence on a given trial exceeds a decision criterion, resulting in categorical decision output (Fig. 2A, right). This framework leads to specific predictions about the impact of variations in detection criterion on neural correlates of decision evidence and decision output. In particular, varying the decision criterion should have no overall effect on the strength of evidence across trials (Fig. 2B), but should affect the evidence strength specifically for detected errors (i.e., hits): If a low (L) criterion is adopted, errors may be detected on the basis of relatively weaker evidence than when a high (H) criterion is used. Thus, to the extent that the Ne/ERN and Pe reflect evidence strength, they should show little overall difference across conditions (i.e., when all errors, both hits and misses, are compared with all correct trials), but should show increased amplitude specifically for hits in the high criterion compared with the low criterion condition. In contrast, neural activity corresponding to decision output should be increased overall in the low criterion condition, because here a higher proportion of errors is detected and signaled, but the activity should be similar when comparing detected errors in the two conditions because of the categorical nature of the decision process (Fig. 2C). In this way, our manipulation of participants' signaling criterion allows us to dissociate the neural correlates of decision evidence and decision output. Note that although we focus here on predictions
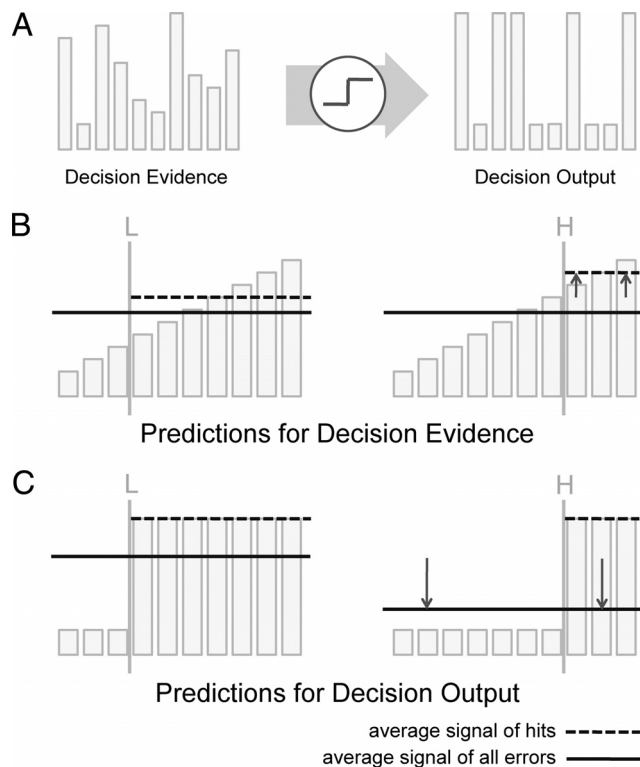
Figure 2. A simple decision model of error detection. **A**, Hypothetical error signals in a sequence of trials. The error detector takes decision evidence as input and determines the decision output by applying a decision criterion. **B**, Effects of criterion shift on decision evidence. Trials are now sorted left to right according to size of error signal, with hypothetical low (L) and high (H) criterion values shown in the left and right panels, respectively. The overall strength of error evidence (solid line) is unchanged, but stronger evidence is required for signaled errors (dashed line) in the high criterion condition. **C**, Effects of criterion shift on decision output. More errors are signaled in the low criterion case, so that overall decision output (solid line) is greater here than in the high criterion condition.

Figure 3. Mean ERP waveforms at electrode FCz for errors and correct responses (upper row) and difference waves for errors minus correct responses (lower row), separately for the low criterion and high criterion conditions. Left column presents waveforms averaged across all trials. Right column presents averaged waveforms including data only from correctly signaled trials (hits and correct rejections). Shaded area indicates the time interval associated with the error negativity (Ne/ERN). Black arrows indicate the latency of the primary task response.
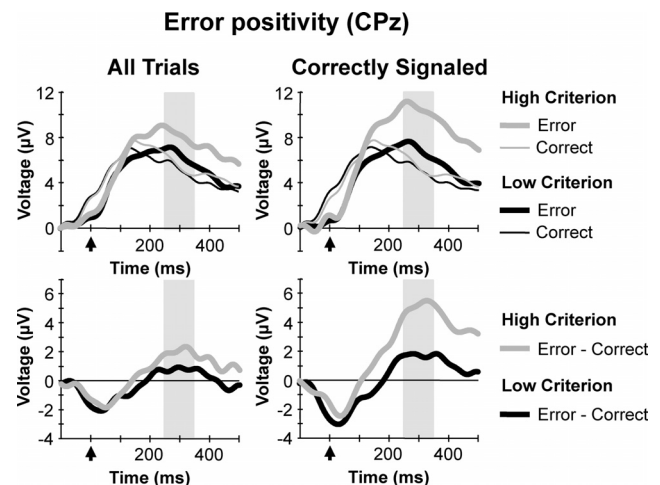
Figure 4. Mean ERP waveforms at electrode CPz for errors and correct responses (top row) and difference waves for errors minus correct responses (bottom row), separately for the low criterion and high criterion conditions. Left column presents waveforms averaged across all trials. Right column presents averaged waveforms including data only from correctly signaled trials (hits and correct rejections). Shaded area indicates the time interval associated with the error positivity (Pe). Black arrows indicate the latency of the primary task response.

for detected errors because our design was optimized for obtaining sufficient numbers of these trials in each condition, corresponding predictions can be derived for undetected errors, which are discussed in the supplemental materials (available at www.jneurosci.org).

To evaluate these predictions, response-locked waveforms for error and correct response trials were computed, first, using all trials (error and correct trials) and, second, using only correctly signaled trials (hits and correct rejections). Figures 3 and 4 show the resulting waveforms at electrodes FCz and CPz. Figure 5 shows the topographies for relevant time intervals. The Ne/ERN is evident as an enhanced negativity following errors compared with correct responses, peaking ~30 ms after the response (Fig. 3) with a frontocentral topography (Fig. 5). The Pe is clearly evident as a subsequent positive deflection that is enhanced on error trials, peaking ~300 ms after the response (Fig. 4) over posterior scalp sites (Fig. 5). As noted above, both components were enhanced on trials in which errors were detected (compare left and right panels of Figs. 3 and 4). However, of central interest are the effects of error signaling criterion on the two components.

As described above, if the Ne/ERN and Pe reflect the output of the error detection process, they should be enhanced overall in the low criterion condition in which error signaling is more frequent. However, contrary to these predictions, Ne/ERN amplitude varied very little across conditions in this comparison ($t < 1$; Fig. 3, left), while Pe amplitude was even somewhat reduced in the low criterion condition (0.97 μV) compared with the high
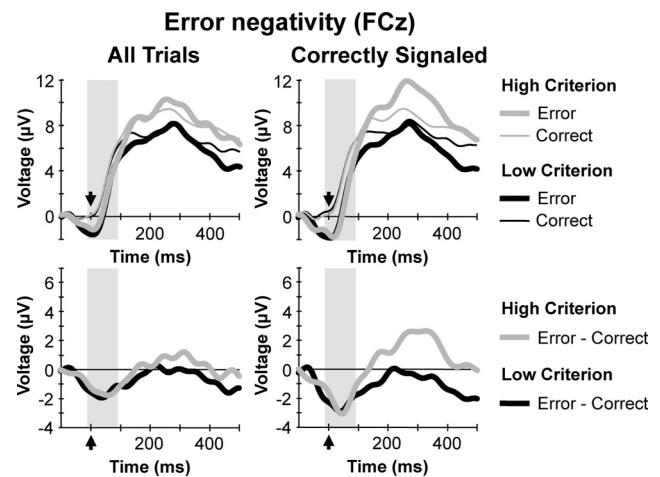
criterion condition (2.46 μV), $t_{(15)} = 3.31$, $p < 0.01$ (Fig. 4, left). These results suggest strongly that the Ne/ERN and Pe do not reflect the output of an error detection system: Neither component showed an increase in amplitude in conditions in which more errors were detected and signaled.

We next considered whether the components might reflect the strength of evidence feeding into decisions about response accuracy: If so, their amplitude should be increased for signaled errors (hits) in the high criterion condition relative to the low criterion condition, because a higher criterion implies that stronger evidence is needed for an error to be detected. As shown in Figure 3 (right), no such effect was apparent for the Ne/ERN ($t < 1$), but the Pe (Fig. 4, right) showed precisely the predicted effect: Its amplitude was markedly increased for hits in the high criterion condition (4.64 μV) compared with the low criterion condition (1.43 μV), $t_{(15)} = 7.40$, $p < 0.001$.
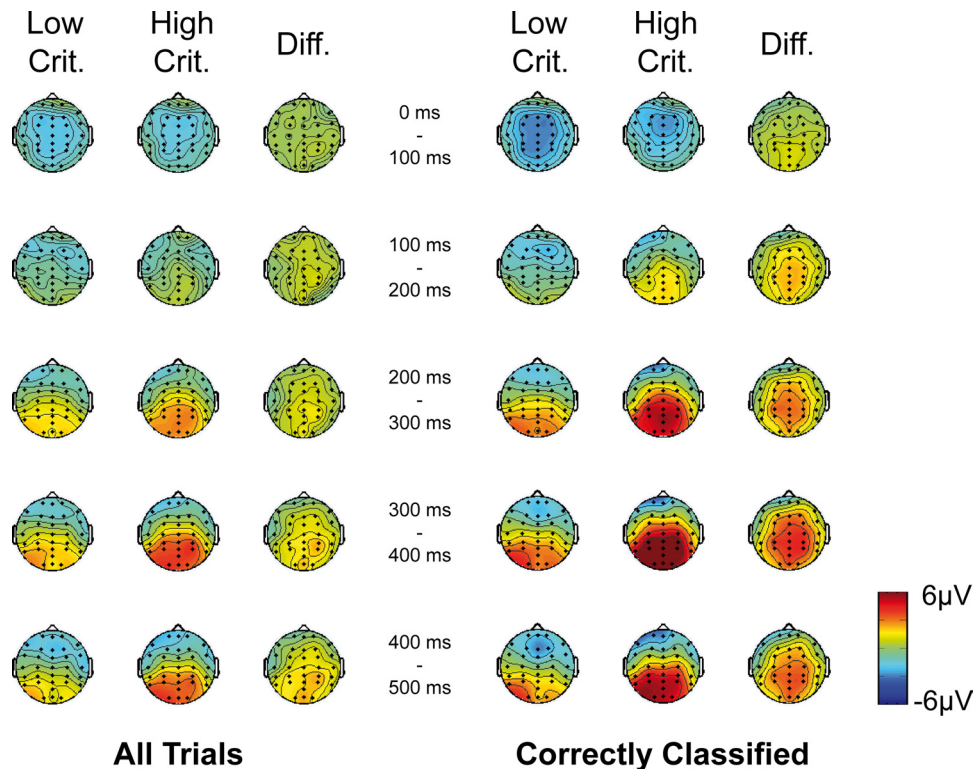
Figure 5. Time course of spatial distribution of the difference between errors and correct responses, separately for all trials and correctly signaled trials for each criterion condition and for the difference between the two criterion conditions. Crit., Criterion.

Two further analyses ruled out possible artifactual explanations of this Pe difference. First, we tested whether the criterion effect for correctly signaled trials exceeded the same effect for all trials, to exclude the possibility that the high criterion condition produced generally increased Pe amplitudes. This was indeed the case, $t_{(15)} = 4.29$, $p < 0.001$. Second, we tested whether the observed effects might be confounded with differences in trial numbers across comparisons (e.g., when comparing hits versus all errors). To this end, we reanalyzed the data after matching trial numbers of each condition. Because matching implies that a randomly chosen subset of trials is analyzed, we repeated the procedure for 1000 different random subsets. The results were always consistent with those of the basic analysis reported above, demonstrating that the reported effects do not reflect confounding effects of differing trial numbers or differing signal-to-noise ratio across conditions. The increased Pe in the high criterion condition is also unlikely to be a direct correlate of the criterion shift, given converging evidence that increasing a decision criterion is achieved by reducing baseline activity in decision-related brain structures (Forstmann et al., 2008; Ivanoff et al., 2008; van Veen et al., 2008; Bogacz et al., 2010). Thus, the results suggest clearly that the Pe reflects the evidence strength that an error has occurred, being greater for hits in conditions in which stronger evidence is required for errors to be detected and signaled.

If our interpretation is correct, then it should be possible to predict the size of participants' behavioral criterion shift based on the criterion effect for the Pe, because the latter effect should provide a neural index of the change in evidence strength across conditions. Consistent with this hypothesis, participants with a larger criterion effect on the Pe showed a larger behavioral criterion shift, as estimated through their signal detection theory parameter, c (Fig. 6). Specifically, to obtain a pure measure of the crucial Pe effect, unconfounded with any overall amplitude dif-
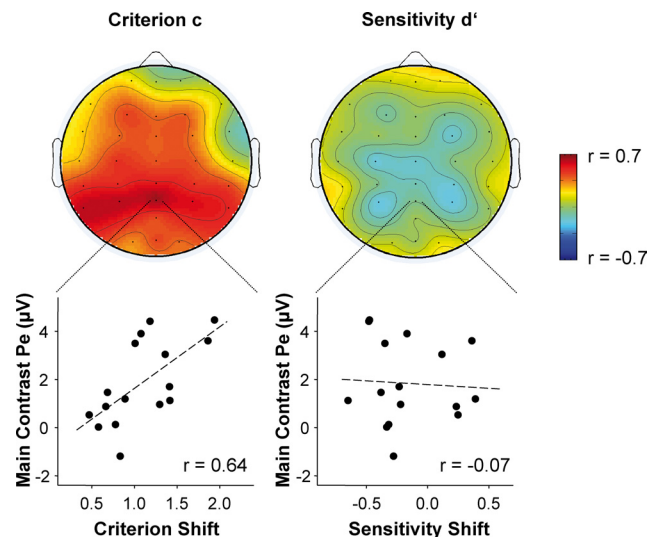

Figure 6. Correlations between the Pe contrast representing the criterion effect for correctly signaled trials minus the criterion effect for all trials, and behavioral estimates of criterion shift (left) and sensitivity shift (right). Scatter plots illustrate correlations at channel Pz.

ference across conditions, we used the Pe contrast representing the criterion effect for correctly signaled errors minus the criterion effect for all errors. The observed correlation between the Pe effect and behavioral criterion shift was robustly observed for posterior scalp locations around channel Pz ($r = 0.64$, $p < 0.01$, Fig. 6), mirroring the posterior topography of the component apparent in the ERP waveform (Fig. 5). In contrast, no significant correlation was apparent between Pe amplitude and detection sensitivity (d') differences between conditions (Fig. 6, right),
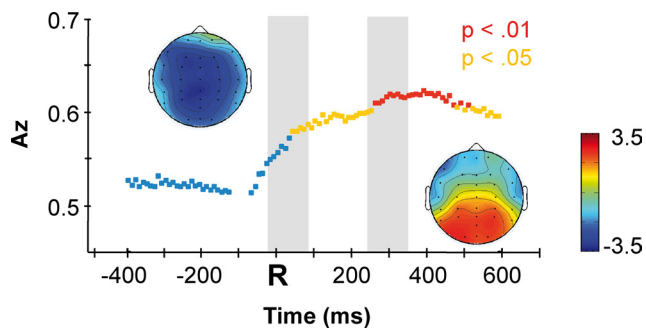
Figure 7. Time course of the extracted component of error-related brain activity, showing sensitivity (Az) of the classifier for discriminating errors and correct responses. Each time point represents the application of the analysis to a moving window of 100 ms width. Time windows close to the baseline interval (−150 to 50 ms) produced implausible values and were omitted. Orange and red points indicate Az classification values significantly above chance (orange: $p < 0.05$, red: $p < 0.01$). The topographies represent the distribution of component activity predicted by the classifier for the marked time periods representing the Ne/ERN (left) and Pe (right).

demonstrating that the Pe effect is specifically related to the criterion shift.

Together, these findings demonstrate that Pe amplitude varied robustly as a function of participants' error signaling criterion—in particular, reflecting the strength of evidence that an error has occurred—whereas the Ne/ERN was relatively insensitive to this manipulation. In the discussion below, we return to the somewhat counterintuitive finding that the Ne/ERN varied so little with error signaling behavior even though its amplitude was significantly greater for signaled than for unsignaled errors. However, we first turn to further exploration of the relationship between the Pe and decision evidence, using multivariate pattern classifier analyses to derive robust estimates of component amplitude on individual trials.

Single-trial analysis: Pe and decision evidence
The preceding analyses suggest that the Pe does not reflect the output of an error detection process, but rather reflects an internal decision variable that conveys information about the evidence on which error detection is based. This hypothesis implies that it should be possible to use the Pe to track the internal processes leading to error detection, and to predict participants' error signaling based on this brain activity that occurred several hundred milliseconds earlier. To evaluate this possibility, we estimated single-trial Pe amplitude using the linear integration classifier method introduced by Parra et al. (2002).

We trained the classifier to differentiate between errors and correct trials, then used its prediction value as a single-trial measure of the error signal (Parra et al., 2002). As an illustration of the ability of the classifier analysis to capture key spatial and temporal features of error-related neural activity, Figure 7 depicts a measure of discrimination performance, Az, for each time point, with orange and red points marking values significantly exceeding chance level. Some Az scores are already significantly above chance in the time period of the Ne/ERN but reach their maximum in the time period of the Pe. The spatial distribution of the discriminating component reveals a frontocentral distribution of weights for the Ne/ERN time period but a more posterior distribution for the Pe time period, demonstrating that the classifier effectively isolates activity corresponding to the components of interest.

Using the derived single-trial measures of the Ne/ERN and Pe time periods, we first investigated whether the amplitude of the two components correlated across trials, separately for each participant. When all trials were entered into the analysis, the resulting correlation coefficients ranged from 0.12 to 0.53 across participants, with a mean of 0.39, which was significantly above zero, $t_{(15)} = 14.3$, $p < 0.001$, and which implies that the Ne/ERN and Pe error signals shared about 15% of their amplitude variance. The correlation decreased when the analysis focused solely on error trials (mean = 0.30, range = 0.16–0.46, $t_{(15)} = 12.3$, $p < 0.001$, shared variance: 9%) or hits (mean = 0.27, range = 0.10–0.45, $t_{(15)} = 9.03$, $p < 0.001$, shared variance: 7%), reflecting the reduction of overall amplitude variance in these trial subsets. Together, these results confirm our earlier conclusion that, although the Ne/ERN and Pe are not completely unrelated, they share relatively little variance and thus seem to convey different types of error information.

We next used the derived error signal estimated from the Pe time period to recover the distribution of activity on error and correct trials, specifically to evaluate whether these distributions can be used to predict behaviorally observed frequencies of error signaling hits and false alarms. To this end, the Pe distributions were used to estimate a hypothetical criterion that best predicted the empirically observed pattern of hit and false alarm rates across the two incentive conditions. More specifically, we applied a criterion K to the distributions and then calculated the expected proportion of false alarms FA (among all correct trials) together with the proportion of hits H (among all errors). By means of exhaustive parameter search, we identified the value for K that minimized the difference between the predicted and the observed values for H/FA, separately for each condition and each participant. Finally, we tested whether the estimated criterion, K, differed significantly across conditions. The results are presented in Figure 8.

As shown in the upper left panel of Figure 8, the Pe distributions for correct and error trials were strongly overlapping. However, when we estimated the hypothetical detection criterion to fit the empirical frequencies of hits and false alarms, we achieved rather good fits (Fig. 8, upper right panel). There were still some deviations—in particular, the neural data ("model") tended to underestimate the observed frequencies ("data") of accurate error signals—but the predicted frequencies successfully captured key trends in the empirical data across the two conditions. Moreover, the estimated criterion, K, was reliably different in the two conditions, $t_{(15)} = 6.99$, $p < 0.001$. These results demonstrate that the recovered error signal in the Pe time period is predictive of subsequent signaling responses and, thus, that the signal provides a valid estimate of decision evidence underlying error detection. Given that our behavioral data can be explained by strongly overlapping distributions of decision evidence on correct and error trials, the overlap seen in Figure 8 appears to reflect the limited sensitivity of error detection in our experiment rather than simply reflecting noise in the EEG error signal. In short, the distribution of neural error signals shown in Figure 8 provides a reasonable approximation of the internal evidence signal on which participants based their error signaling responses.

Figure 8 also illustrates the utility of the logistic regression classifier. The bottom presents the results of analyses in which single-trial Pe amplitudes were estimated simply using the ERP waveform at CPz (the electrode at which the Pe is maximal) from 250 to 350 ms post-response. Using this measure, the distributions for errors and correct trials are nearly completely overlapping: Variability of the signal in each distribution (mean SD = 9.60 μV and 9.70 μV, for correct and error trials, respectively) is much larger than the difference in distribution means (1.99 μV).

Moreover, using these distributions, the best-fitting signaling criterion did not differ for the low and high criterion conditions, $t_{(15)} = 1.00$, $p = 0.33$, and there were strong deviations of observed and predicted frequencies for both conditions. Overall, therefore, it is evident that raw ERP voltages do not provide a suitably robust measure of underlying neural activity to predict overt performance on single-trials, in contrast to the values derived from the linear classifier analysis described above.

Our final analysis assessed whether classifier-derived Pe values robustly predict individual signaling responses. As described above, the classifier was trained to discriminate errors from correct responses; it was not trained to discriminate hits from misses or correct rejections from false alarms. It is nevertheless possible to assess whether single-trial Pe values effectively distinguish these trial types. An analysis of this effect produced a classification Az of 0.62 for the discrimination of hits vs misses, a level robustly above chance, $p < 0.01$, as revealed by a permutation test (1000 permutations; critical values: 0.582 for $p < 0.05$, 0.615 for $p < 0.01$). Classifier Pe values discriminated false alarms from correct rejections at a similar level (Az = 0.61), and again robustly above chance, $p < 0.05$ (1000 permutations; critical values: 0.586 for $p < 0.05$, 0.619 for $p < 0.01$). Thus, classifier Pe values can be used to predict individual signaling responses, albeit imperfectly.

Figure 9 presents the data from individual error trials across all participants to illustrate the effectiveness of single-trial classification and the level of signal variability. For each participant, the error data are sorted according to the classifier-derived Pe value, shaded according to whether they were accurately signaled (light gray) or missed (dark gray), and plotted against the estimated criterion values from the preceding analysis. For most participants, misses were clearly more frequent on trials with smaller error signals (i.e., the dark gray bars cluster to the left of the plots), demonstrating that recovered Pe values provide a valid predictor of the subsequent signaling response on individual trials as well as on an aggregate level across trials.

Discussion

The present study provides new insight into the functional significance of error-related EEG activity by treating performance monitoring as a decision-making process. Within this framework, we asked whether the Ne/ERN and Pe reflect the strength of evidence that an error has occurred, or the output of the error detection process. If the latter were true, then the Ne/ERN and Pe should be large in conditions in which participants frequently detected and signaled their errors. However, no such effect was apparent for the Ne/ERN, and Pe amplitude was even somewhat reduced when participants were encouraged (via performance incentives) to signal their errors frequently. Instead, Pe amplitude varied with the strength of evidence that an error had occurred, being larger for
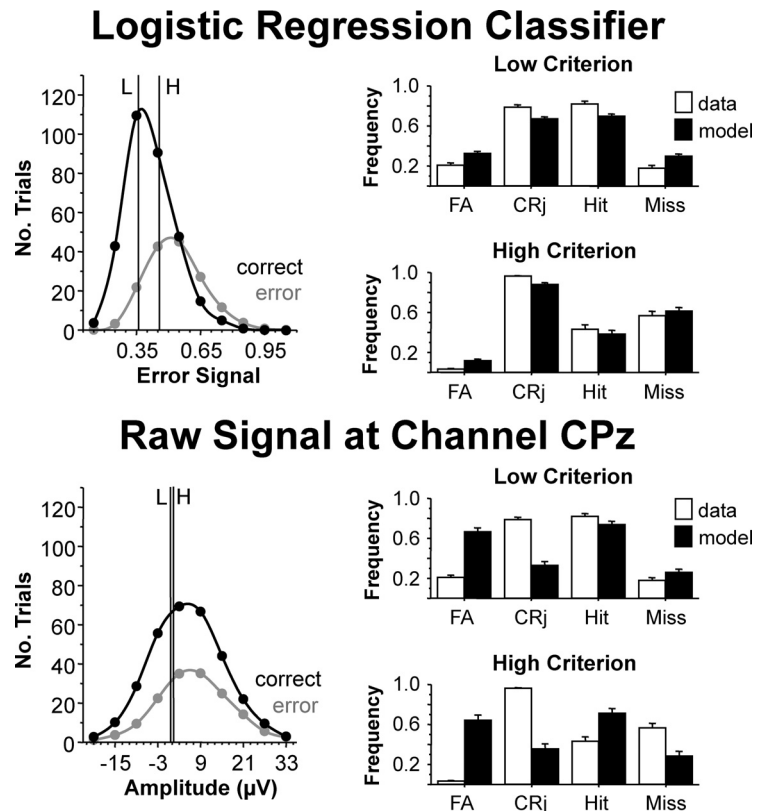

Figure 8. Recovered distribution of error signals estimated by the prediction value of the classifier (top row) and raw ERP voltage at channel CPz (bottom row) for each trial. Left column, Separate distributions for correct trials and errors were constructed. Vertical lines indicate the estimated criterion values for the low criterion (L) and high criterion (H) conditions. Right column, Empirical ("data") and predicted ("model") frequencies of false alarms (FA), correct rejections (CRj), hits, and misses, for each condition.

detected errors when participants adopted a strict signaling criterion (i.e., when strong evidence was required for error detection, and relatively few errors were signaled). No such effect was observed for the Ne/ERN. Extending these analyses, we found that participants' error detection criterion—estimated using signal detection theory analysis—could be predicted from variations in Pe amplitude, suggesting further that this component reflects the internal evidence that an error has been committed.

The observed dissociation between the Ne/ERN and Pe is consistent with the proposal that these two components reflect at least partially dissociable aspects of performance monitoring (Overbeek et al., 2005). In the present data, Ne/ERN and Pe amplitude correlated across trials and both were larger for detected than undetected errors (cf. Scheffers and Coles, 2000; Maier et al., 2008), but only Pe amplitude varied across conditions in a manner predictive of changes in participants' signaling behavior. Thus, consistent with previous work (Nieuwenhuis et al., 2001; Endrass et al., 2005, 2007; O'Connell et al., 2007; Shalgi et al., 2009), our findings suggest that the Pe correlates more closely with subjective judgments of response accuracy than does the Ne/ERN.

It initially seems puzzling that Ne/ERN amplitude varied with some aspects of overt error detection (being larger for detected errors) but not others (being insensitive to changes in detection criterion). However, both results are consistent with the hypothesis that the Ne/ERN reflects intrinsic features of task performance—such as the occurrence of response conflict (Yeung et al., 2004) or the probability of errors (Holroyd and Coles, 2002)—
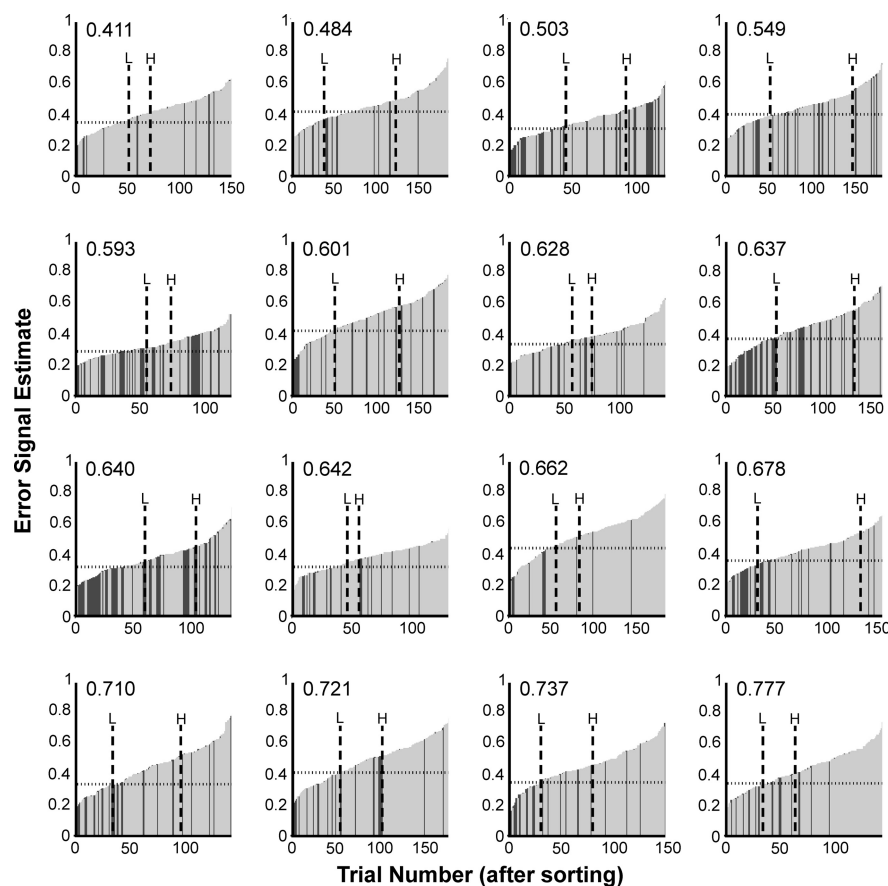
Figure 9. Classifier-based estimates of Pe amplitude for error trials sorted by value (signal strength), separately for each participant. Light gray indicates hits, dark gray indicates misses. The number in the upper left corner refers to the discrimination sensitivity between hits and misses (Az). Vertical lines are behaviorally estimated low (L) and high (H) criterion values. Horizontal dashed lines indicate mean of error signal.

P3 as a neural response to a salient event (Ridderinkhof et al., 2009), it has remained unclear which aspect of error processing is indexed by this component. For example, a recent review concluded that "it remains to be clarified whether the Pe is the *expression* of error awareness, or reflects the processes that *lead to* error awareness" (Ridderinkhof et al., 2009, p. 536, italics original). The present findings demonstrate that the Pe reflects an early stage of error processing, before categorical decisions about response accuracy. Thus, Pe amplitude did not vary simply with the rate of error signaling, and indeed was somewhat larger in the high criterion condition in which fewer errors were detected. The increase in Pe amplitude in the high criterion condition became marked when our analyses focused on detected errors. These combined findings are best explained by assuming that only trials with strong evidence are signaled as errors when a high criterion is applied, and that the Pe reflects the strength of this evidence. This interpretation receives further support from analyses showing that the criterion effect for the Pe was predictive of participants' behavioral criterion shifts as estimated using signal detection theory analysis.

To explore further the relation between the Pe and decisions underlying error detection, we examined whether its amplitude is predictive of subsequent behavioral error signaling responses both at an aggregate level (in terms of hit and false alarm rates) and at a single-trial level (in terms of individual signaling responses). To this end, the distribution of error-related EEG activity across trials was computed using logistic regression classification. Classification performance was maximal in the latency range of the Pe. Critically, EEG activity at this latency predicted the pattern of hit rates and false alarm rates in the two conditions as well as the criterion shift. In this way, our analyses suggest that the Pe provides a robust index of the internal weight of evidence that an error has occurred: The distribution of Pe amplitude across trials (Figs. 8, 9) may therefore be the neural basis of the hypothesized variable-strength error signal (Fig. 2). Consistent with this interpretation, our classifier-based Pe measure predicted participants' signaling responses on single-trials at a level robustly above chance. Together, these results suggest strongly that the Pe conveys probabilistic information about the occurrence of an error, information that may subsequently lead to overt judgments about response accuracy and necessary remedial actions.

A critical feature of the present approach is the application of established principles from decision making research (Green and Swets, 1966; Ratcliff and Rouder, 1998). We thus characterized error detection as involving the accumulation of evidence for an error that is compared against a decision criterion. This interpretation raises the intriguing possibility that performance monitoring decisions might rely on similar neural mechanisms to other, well characterized decision processes such as those involved in perceptual categorization (Gold and Shadlen, 2007; Heekeren et

rather than providing a direct index of error processing. According to this hypothesis, Ne/ERN amplitude should be determined primarily by variations in primary task performance rather than variations in error signaling. Critically for this interpretation, task performance differed significantly for detected and undetected errors, being consistently faster for the former. Thus, the Ne/ERN increase for detected errors may not reflect its direct role in error processing, but might instead be a by-product of the fact that detected errors tend to occur when fast guess responses are subsequently corrected (cf. Scheffers and Coles, 2000), resulting in high levels of conflict. This interpretation is consistent with evidence from the antisaccade task that Ne/ERN amplitude is similar for detected and undetected errors that are always corrected (Nieuwenhuis et al., 2001), although in some studies this relationship is less clear (Endrass et al., 2007). In contrast, primary task performance was very little affected by variations in error signaling criterion. Ne/ERN amplitude correspondingly varied little across conditions, despite marked differences in error detection performance. Together, these findings provide further evidence that the Ne/ERN does not directly index error monitoring processes, contrary to early theoretical accounts of this component (Falkenstein et al., 1990; Gehring et al., 1993).

The Pe appears much more directly related to explicit error detection than the Ne/ERN. While this conclusion converges with other recent evidence (Nieuwenhuis et al., 2001; Endrass et al., 2005, 2007; O'Connell et al., 2007; Shalgi et al., 2009), and is broadly consistent with the suggestion that the Pe resembles the

al., 2008) that have been analyzed using similar methods to those adopted here (Philiastides and Sajda, 2006, 2007; Philiastides et al., 2006). This hypothesis converges with recent suggestions that the Pe may share neural generators with the stimulus-related P3 component, and that both reflect the conscious processing of motivationally salient events (Ridderinkhof et al., 2009). Moreover, neurons in the ventral premotor cortex of monkeys have been identified that support both perceptual decisions and error monitoring (Pardo-Vazquez et al., 2008). On this view, the benefits of treating performance monitoring as a decision process may be more than simply methodological: To the degree that there are shared neural mechanisms for accumulating and evaluating evidence about external (sensory) events and internal (monitoring) processes, understanding of performance monitoring will be deepened by further investigation of these shared processes.

Future research might therefore build on the present approach by further specifying the neural correlates of decision evidence and identifying correlates of decision output in performance monitoring. For example, the present approach might be extended using functional magnetic resonance imaging to identify stages of error processing implemented in specific networks previously shown to be differentially active for detected and undetected errors (Klein et al., 2007; Ullsperger et al., 2010), or using formal models to guide single-trial analysis (Cavanagh et al., 2010; Philiastides et al., 2010). Such issues might fruitfully be addressed using extensions of the present methods, in which experimental manipulations were targeted directly at the error detection process—through varying incentives for error signaling—rather than influencing error detection indirectly via modulations of primary task performance (cf. Yeung et al., 2007). This approach allowed us to dissociate neural activity that varied with participants' error signaling judgments (and therefore varied across conditions) from activity more closely tied to the primary behavioral task (which therefore varied relatively little). Our primary task was in turn designed to allow precise control over the absolute number of errors as well as the rate of undetected errors. Using these methods, the present study demonstrates that principles and methods from decision making research may shed new light on the mechanisms underlying performance monitoring. In particular, our findings suggest that the Pe reflects an internal evidence signal that an error has occurred, and as such represents an early stage of error detection and compensation in the service of optimizing task performance.

References

Bogacz R, Wagenmakers EJ, Forstmann BU, Nieuwenhuis S (2010) The neural basis of the speed-accuracy tradeoff. Trends Neurosci 33:10–16.

Cavanagh JF, Frank MJ, Klein TJ, Allen JJ (2010) Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. Neuroimage 49:3198–3209.

Endrass T, Franke C, Kathmann N (2005) Error awareness in a saccade countermanding task. J Psychophysiol 19:275–280.

Endrass T, Reuter B, Kathmann N (2007) ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. Eur J Neurosci 26:1714–1720.

Falkenstein M, Hohnsbein J, Hoormann J, Blanke L (1990) Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In: Psychophysiological brain research (Brunia CHM, Gaillard AWK, Kok A, eds), pp 192–195. Tilburg, The Netherlands: Tilburg UP.

Forstmann BU, Dutilh G, Brown S, Neumann J, von Cramon DY, Ridderinkhof KR, Wagenmakers EJ (2008) Striatum and pre-SMA facilitate decision making under time pressure. Proc Natl Acad Sci U S A 105:17538–17542.

Gehring WJ, Goss B, Coles MG, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. Psychol Sci 4:385–390.

Gold JI, Shadlen MN (2007) The neural basis of decision making. Annu Rev Neurosci 30:535–574.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.

Hajcak G, Moser JS, Yeung N, Simons RF (2005) On the ERN and the significance of errors. Psychophysiology 42:151–160.

Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. Nat Rev Neurosci 9:467–479.

Holroyd CB, Coles MG (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 109:679–709.

Ivanoff J, Branning P, Marois R (2008) fMRI evidence for a dual process account of the speed-accuracy tradeoff in decision-making. PLoS One 3:e2635.

Kaernbach C (1991) Simple adaptive testing with a weighted up-down method. Percept Psychophys 49:227–229.

Kiehl KA, Liddle PF, Hopfinger JB (2000) Error processing and the rostral anterior cingulate: an event-related fMRI study. Psychophysiology 37:216–223.

Klein TA, Endrass T, Kathmann N, Neumann J, von Cramon DY, Ullsperger M (2007) Neural correlates of error awareness. Neuroimage 34:1774–1781.

Macmillan NA, Creelman CD (1991) Detection theory: a user's guide. Cambridge, UK: Cambridge UP.

Maier M, Steinhauser M, Hübner R (2008) Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. J Cogn Neurosci 20:2263–2273.

Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GP, Kok A (2001) Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. Psychophysiology 38:752–760.

O'Connell RG, Dockree PM, Bellgrove MA, Kelly SP, Hester R, Garavan H, Robertson IH, Foxe JJ (2007) The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. Eur J Neurosci 25:2571–2579.

Overbeek TJM, Nieuwenhuis S, Ridderinkhof KR (2005) Dissociable components of error processing: on the functional significance of the Pe vis-à-vis the ERN/Ne. J Psychophysiol 19:319–329.

Pailing PE, Segalowitz SJ (2004) The effects of uncertainty in error monitoring on associated ERPs. Brain Cogn 56:215–233.

Pardo-Vazquez JL, Leboran V, Acuña C (2008) Neural correlates of decisions and their outcomes in the ventral premotor cortex. J Neurosci 28:12396–12408.

Parra L, Alvino C, Tang A, Pearlmutter B, Yeung N, Osman A, Sajda P (2002) Linear spatial integration for single-trial detection in encephalography. Neuroimage 17:223–230.

Parra LC, Spence CD, Gerson AD, Sajda P (2005) Recipes for the linear analysis of EEG. Neuroimage 28:326–341.

Philiastides MG, Sajda P (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 16:509–518.

Philiastides MG, Sajda P (2007) EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. J Neurosci 27:13082–13091.

Philiastides MG, Ratcliff R, Sajda P (2006) Neural representation of task difficulty and decision-making during perceptual categorization: a timing diagram. J Neurosci 26:8965–8975.

Philiastides MG, Biele G, Vavatzanidis N, Kazzer P, Heekeren HR (2010) Temporal dynamics of prediction error processing during reward-based decision making. Neuroimage 53:221–232.

Rabbitt PM (1968) Three kinds of error-signalling responses in a serial choice task. Q J Exp Psychol 20:179–188.

Ratcliff R, Rouder JN (1998) Modeling response times for two-choice decisions. Psychol Sci 9:347–356.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. Science 306:443–447.

Ridderinkhof KR, Ramautar JR, Wijnen JG (2009) To PE or not to PE: a P3-like ERP component reflecting the processing of response errors. Psychophysiology 46:531–538.

Scheffers MK, Coles MG (2000) Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. J Exp Psychol Hum Percept Perform 26:141–151.

Semlitsch HV, Anderer P, Schuster P, Presslich O (1986) A solution for

reliable and valid reduction of ocular artifacts, applied to the P300 ERP. Psychophysiology 23:695–703.

Shalgi S, Barkan I, Deouell LY (2009) On the positive side of error processing: error-awareness positivity revisited. Eur J Neurosci 29:1522–1532.

Steinhauser M, Maier M, Hübner R (2008) Modeling behavioral measures of error detection in choice tasks: response monitoring versus conflict monitoring. J Exp Psychol Hum Percept Perform 34:158–176.

Ullsperger M, von Cramon DY (2001) Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. Neuroimage 14:1387–1401.

Ullsperger M, Harsay HA, Wessel JR, Ridderinkhof KR (2010) Conscious

perception of errors and its relation to the anterior insula. Brain Struct Funct 214:629–643.

van Veen V, Carter CS (2002) The timing of action-monitoring processes in the anterior cingulate cortex. J Cogn Neurosci 14:593–602.

van Veen V, Krug MK, Carter CS (2008) The neural and computational basis of controlled speed-accuracy tradeoff during task performance. J Cogn Neurosci 20:1952–1965.

Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev 111:939–959.

Yeung N, Ralph J, Nieuwenhuis S (2007) Drink alcohol and dim the lights: the impact of cognitive deficits on medial frontal cortex function. Cogn Affect Behav Neurosci 7:347–355.