Journal of Experimental Psychology:
Human Perception and Performance
2008, Vol. 34, No. 1, 158–176

Copyright 2008 by the American Psychological Association
0096-1523/08/$12.00 DOI: 10.1037/0096-1523.34.1.158

Modeling Behavioral Measures of Error Detection in Choice Tasks: Response Monitoring Versus Conflict Monitoring

Marco Steinhauser, Martin Maier, and Ronald Hübner
Universität Konstanz

The present study investigated the mechanisms underlying error detection in the error signaling response. The authors tested between a response monitoring account and a conflict monitoring account. By implementing each account within the neural network model of N. Yeung, M. M. Botvinick, and J. D. Cohen (2004), they demonstrated that both accounts make different predictions as to how error signaling performance is influenced by varying the participants' response criterion. These predictions were tested in an experiment using the Eriksen-flanker task. The qualitative pattern as well as a model fit favored the response monitoring account, which states that error detection is mediated by detecting internal error corrections.

Keywords: error detection, error correction, response conflict, connectionist modeling, Eriksen-flanker task

The ability to detect errors is crucial for the adaptability of the mental system. It supports the optimization of strategies (Laming, 1979; Ridderinkhof, 2002) as well as the acquisition of skills (Holroyd & Coles, 2002). Therefore, the investigation of error detection contributes substantially to an understanding of human cognition. Whereas early efforts almost exclusively focused on behavioral indicators of error detection (e.g., Rabbitt, 1966a, 1966b), recently, psychophysiological measures such as event-related potentials (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990; Gehring, Goss, Coles, Meyer, & Donchin, 1993) have been the main interest.

In the present article, we demonstrate that behavioral data can still be valuable for gaining insight into the nature of error processing. Our goal was to investigate the mechanism underlying error signaling, which is a classical behavioral measure of error detection (e.g., Rabbitt, 1968), by comparing the predictions of two prominent models in this area: The idea that error detection consists in monitoring whether an overt response is internally corrected (e.g., Rabbitt, Cumming, & Vyas, 1978), which we call the *response monitoring account*, and the idea that this is achieved by monitoring response conflict, which is called the *conflict monitoring account* (Yeung, Botvinick, & Cohen, 2004).

To attain this objective, we used the connectionist model of Yeung et al. (2004), which can simulate psychophysiological as well as behavioral measures. Although this model realizes a conflict monitoring account, it can also be used to implement the basic ideas underlying a response monitoring account of error detection. Our method was to simulate both accounts and to test which one provided a better fit to the behavioral measures of error detection. Before we report the experimental data and the modeling results, we give a short overview of the theories involved and relevant empirical measures.

Response Monitoring

In choice tasks, a stimulus usually has to be classified by producing a speeded response according to a prespecified rule. The involved response selection process is often thought to proceed by evidence accumulation for each possible response until a certain response criterion is exceeded (e.g., Ratcliff & Rouder, 1998). In most cases, this process should select the response that accumulates evidence at the highest rate, and that is usually correct. However, because the process is noisy, sometimes a wrong response is selected. This raises the question of how such errors are detected by the system.

A possible answer is provided by what we call the response monitoring (RM) account of error detection. In the context of an evidence accumulation model, the idea of RM implies that a mechanism registers the resulting response whenever the accumulated evidence has exceeded a criterion. Moreover, after a response has been selected and produced, the accumulation of evidence continues. Consequently, further evidence could lead to the selection of a second response. If this occurs, the monitoring mechanism compares the second response with the first one, and if there is a discrepancy, it concludes that the first response was an error. Such a mechanism enables the system to detect errors, given that the later response is more reliable than the earlier one. Conceived in this way, error detection is equivalent to the detection of an *internal correction response*. Interestingly, whenever an error has been detected, the RM system also knows the identity of the correct response, because this is represented by the correction response.

Marco Steinhauser, Martin Maier, and Ronald Hübner, Fachbereich Psychologie, Universität Konstanz, Konstanz, Germany.

This research was supported by Deutsche Forschungsgemeinschaft Grant Hu 432/8 to Ronald Hübner. We thank Markus Ullsperger and Nick Yeung for valuable comments on a previous version of this article.

Correspondence concerning this article should be addressed to Marco Steinhauser, Fachbereich Psychologie, Universität Konstanz, Fach D29, D-78457, Konstanz, Germany. E-mail: marco.steinhauser@uni-konstanz.de

The central ideas of this account have been formulated before. Similar assumptions underlie, for instance, the committee decision model by Rabbitt and colleagues (Rabbitt et al., 1978; Rabbitt & Vyas, 1981), which was initially developed to explain the ability to correct and detect errors very quickly. In those studies, participants either had to correct errors immediately or indicate a detected error (e.g., by pressing a neutral response key). Both types of responses were rather fast. For instance, some error corrections occurred less than 40 ms after the erroneous response.

More recently, Falkenstein et al. (1990) and Gehring et al. (1993) independently discovered that errors are accompanied by a negative deflection in the response-locked event-related potential on frontocentral channels peaking about 100 ms after the erroneous response. This phenomenon, the *error negativity* (Ne; Falkenstein et al., 1990) or *error-related negativity* (ERN; Gehring et al., 1993), was initially taken as evidence for an RM account of error detection. More specifically, both groups of authors suggested that the Ne/ERN is related to a comparator process that compares the intended correct response with the actual one. This idea received further support from the observation that the amplitude of the Ne/ERN is related to the discrepancy between the erroneous and the correct response (Bernstein, Scheffers, & Coles, 1995; Falkenstein, Hohnsbein, & Hoormann, 1995).

Conflict Monitoring

An alternative account of error detection has been proposed by Yeung et al. (2004) within the framework of the conflict monitoring (CM) theory (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Carter et al., 1998). The CM theory assumes that the registration of conflicts between competing responses is an important mechanism for action evaluation. A response conflict is present whenever two or more responses are activated concurrently, or, in other words, when strong evidence has been accumulated for multiple responses at the same time. Botvinick et al. (2001) suggested that the detection of response conflicts generally supports the flexible adaptation of behavior in a variety of tasks. Already Carter et al. (1998) had argued that the Ne/ERN is not related to error processing per se but, rather, reflects the amount of a response conflict, which is generally high on error trials (but see also Luu, Flaisch, & Tucker, 2000).

Recently, these ideas have been elaborated by Yeung et al. (2004). In their neural network model, the Ne/ERN reflects the response conflict that emerges after an erroneous response. Their model shares a central idea with the RM account: After a response, stimulus processing continues, which, in case of an error, leads to the activation of the correct response. Crucially for their model, however, this implies that the correct and the erroneous response are activated simultaneously for a short period after the error. The resulting response conflict is reflected by the Ne/ERN. In this way, Yeung et al. explained some findings from the Ne/ERN literature, which were thought to be incompatible with the CM account. Although their model was mainly constructed to account for the Ne/ERN, they additionally developed a CM account of error detection. In their model, an error is detected whenever the accumulated response conflict after the first response exceeds a threshold.

Taken together, the RM and CM accounts share the idea that stimulus evaluation continues after the first response. However, they differ with respect to how errors are detected. Whereas the RM account assumes that error detection operates by the detection of internal correction responses, the CM account implies that it is the accumulated posterror response conflict that indicates an error. The aim of the present study was to distinguish between these two theories empirically.

One approach could be to examine the source of the Ne/ERN, because both accounts differ with respect to the interpretation of this component. However, the question of which mechanism underlies the Ne/ERN and that about the nature of error detection are not necessarily identical. It is possible that the Ne/ERN reflects a response conflict, although error detection is accomplished by a response monitoring process. Therefore, it is helpful to address both questions separately. But what would be an alternative measure of error detection? We propose that behavioral measures of error detection can serve this purpose. As is shown in the following, the two accounts of error detection differ in their assumptions about what is reflected by the behavioral measures.

Behavioral Measures of Error Detection

As already mentioned, Rabbitt and colleagues (Rabbitt et al., 1978; Rabbitt & Vyas, 1981) considered two behavioral measures of error detection. Their participants had to either correct errors immediately by giving the correct response (*error correction response*; ECR) or indicate a detected error, for instance by pressing a neutral key or by simultaneously pressing all response keys (*error signaling response*; ESR). The duration of ECRs and ESRs was measured as the time elapsed between the erroneous response and the respective detection response.

In several studies, it has been shown that a number of variables affect the ECRs or ESRs. For instance, stimulus masking (Rabbitt & Vyas, 1981) and increasing the number of response alternatives impairs ECRs (Rabbitt & Rodgers, 1977), whereas stimulus–response compatibility affects ECRs (Rabbitt & Phillips, 1967) as well as ESRs (Rabbitt, 1967). Interestingly, Rabbitt (1990, 2002) compared both measures and found that ECRs are faster and occur more frequently than ESRs. He concluded that the former are more automatic than the latter. This conclusion was further supported by the fact that error corrections sometimes occurred spontaneously even though they were not required (Fiehler, Ullsperger, & Von Cramon, 2005; Rabbitt & Rodgers, 1977). Furthermore, the presentation of a new stimulus immediately after an error interferes with ESRs but not with ECRs (Rabbitt, 2002). We discuss the reasons for this difference later.

Important at this point is the fact that the two measures differ in another respect. From a theoretical view, ESRs and ECRs could be based on different information. Whereas an ESR merely requires that an error is detected, for an ECR the system requires also a representation of the correct response. In the following, we argue that this fact can be used to distinguish between the RM and CM accounts of error detection, because they differ in their interpretation of the relation between error detection and error correction.

Basically, the RM account assumes that ESRs and ECRs are based on the same internal error correction process. If an internal correction response occurs, then it can either be used to overtly correct the error or simply to signal the error. In contrast, the CM account assumes that ESRs rely on the detection of a response conflict. Because the amount of conflict depends only on the simultaneous activation of two or more responses, an ESR does not necessarily require information about the correct response. Only if an ECR is required does the system needs information about the correct response, which results from an internal correction response. Therefore, it should be possible to distinguish between the two accounts by manipulating the internal correction response. If this manipulation affects both ESRs and ECRs in the same way, this would support the RM account. However, if it only affects ECRs, the CM account would be supported. The question is how such a manipulation can be accomplished.

A variable that should directly affect the internal correction response is the *response criterion*. It is reasonable to assume that this criterion influences not only the initial response but also the correction response. The more conservative the criterion, the slower the initial response should be. But at the same time, the correction response should also be slower. This fact can be used to distinguish between RM and CM. According to the RM account, the variation of the response criterion should affect ECRs and ESRs in the same way, because both rely on the same internal correction response. In contrast, the CM account does not make such a prediction. Although the response criterion should affect ECRs, the ESR performance should be unaffected, because it relies on the monitoring of a response conflict rather than on internal correction.

However, the assumption that the CM account predicts no criterion effect at all on ESR performance might be too strong. Indeed, there is some evidence suggesting that the response criterion affects the response conflict and, in this way, also the efficiency of CM-based error detection. For instance, Yeung et al. (2004) demonstrated with their CM model that the response conflict as well as the frequency of error detection is altered when the response criterion and an attention parameter are manipulated simultaneously. Unfortunately, they did not examine whether the response criterion alone could be responsible for this effect or whether there was also an effect on the detection latency. Nevertheless, Yeung et al.'s finding suggests that the CM account could also predict a criterion effect on ESR performance. As a consequence, if it turned out that varying the response criterion affected ESRs and ECRs in the same way, as predicted by the RM account, we would not know whether this happened because both were based on an internal correction or because the response criterion affected ESRs indirectly through the response conflict. To deal with this problem, we first had to verify that the two accounts indeed make distinguishable predictions. Fortunately, this could be accomplished by combining empirical testing with computational modeling.

The Present Approach

The RM account predicts that the response criterion affects ECR and ESR performance in a similar way, because both rely on the same internal correction response. However, it is open whether the same pattern is predicted by the CM account. Therefore, we implemented both accounts as extended versions of the neural network model of Yeung et al. (2004). In this way, we could derive exact quantitative predictions for each account.

We proceeded in three steps. First, we compared both models in a series of simulations. This should demonstrate that the two accounts really make differential predictions. In addition, the simulations should uncover the mechanisms responsible for the criterion effects on error processing. The derived predictions were then tested in an experiment. These two steps, however, were not sufficient to definitely differentiate between the models. Therefore, in a final step, we fitted both models to our data to see which is more appropriate.

Because our approach was strongly based on Yeung et al.'s (2004) model, we used an experimental paradigm similar to the one that these authors used in their simulations. Yeung et al. simulated an Eriksen-flanker task in which a target letter that was surrounded by several identical distractor letters has to be classified (Eriksen & Eriksen, 1974). The distractors could be linked either to the same (congruent) or to the alternative (incongruent) response. As stimuli, Yeung et al. used two letters, *H* and *S*, and two corresponding responses.

Different from Yeung et al's (2004) procedure, we decided to use a three-response paradigm. This is crucial because only with more than two possible responses do reliable ECRs require that the system derives the correct response. In a two-response paradigm, it is sufficient to detect an error and produce the alternative response. In the latter case, similarities between ESR and ECR performance could be explained by assuming that both rely on the same (CM-based) error detection mechanism. Only by using three responses could we be sure that similar results for ECRs and ESRs indicated that both rely on internal error correction. In addition, we applied a larger stimulus set and used neutral stimuli instead of congruent ones.

The Models

The original model of Yeung et al. (2004) consists of two parts: a task module, which is based on earlier implementations of the Eriksen-flanker task (Cohen, Servan-Schreiber, & McClelland, 1992; Servan-Schreiber, Bruno, Carter, & Cohen, 1998; Servan-Schreiber, Carter, Bruno, & Cohen, 1998; Spencer & Coles, 1999), and a CM module, which registers response conflicts in the task module, as proposed by Botvinick et al. (2001). For the present purpose, we adapted the task module to our paradigm by adding additional response and stimulus units. On the basis of this modification, we constructed two extensions, one according to the CM account and the other according to the RM account.

In the following, we focus on those aspects that are important for our objective. A formal description of the model can be found in Appendix A. We first present the details of the task module. Then, we describe how we constructed the RM and CM models. Finally, we present simulated results that served to derive predictions for the two accounts with respect to how the response criterion affects the behavioral measures of error detection.

The Task Module

The task module implements the Eriksen-flanker task as a simple three-layer neural network. First of all, there is a set of stimulus units that is connected to a set of response units. In our version (see Figure 1), the stimulus layer consists of one unit for each possible stimulus at each of the three possible display positions (left, center, right). Each letter unit is unidirectionally connected to one of the three response units in the response layer. Units representing neutral stimuli are not connected to any re-

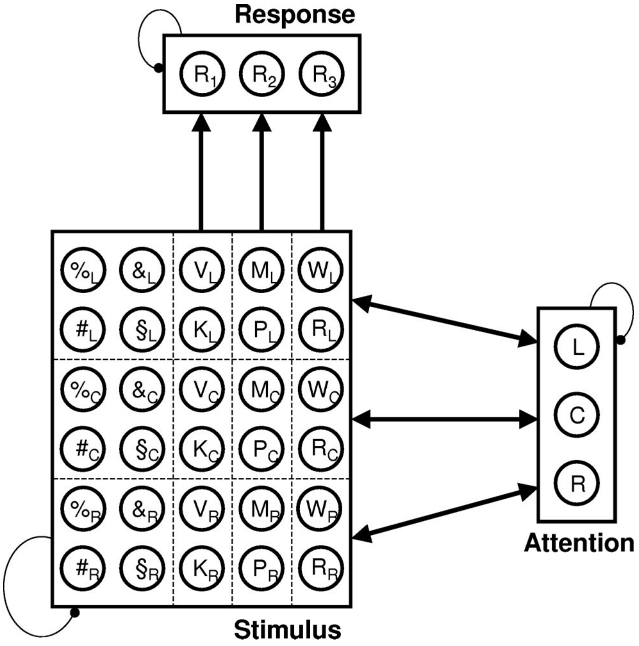

Figure 1. Modified version of the task module of Yeung et al. (2004). Each unit in the stimulus layer represents a stimulus (%, &, #, §, V, K, M, P, W, R) on a specific display position (indices L, C, and R). Unidirectional links connect each stimulus unit representing a letter with its corresponding response unit in the response layer (e.g., V_L, V_C, and V_R are connected to R_1), whereas no such connections exist for the stimulus units representing the neutral symbols. Bidirectional links connect each stimulus unit to its corresponding position unit in the attention layer (e.g., $\%_L$ and V_L are connected to L). L = left; C = center; R = right.

sponse. Finally, there is a third attention layer consisting of one unit for each position. By means of bidirectional connections, each input layer unit is connected to its corresponding position unit in the attention layer. Within each layer, the units are connected by inhibitory associations.

At the beginning of each trial, a stimulus was presented to the model by activating its corresponding pattern in the stimulus layer. For instance, the stimulus *VPV* implied that the left and right V units (V_L, V_R) and the central P unit (P_C) were activated. In addition, an attentional set was realized by activating the center unit in the attention layer more strongly than the lateral units.[1] Because the attention units are connected to the stimulus units, the activation of the target in the stimulus layer became more pronounced after some cycles than the activations of the flanking stimuli.

The feedforward connections from the stimulus to the response layer led to an accumulation of activation in the response units. A response was selected as soon as the activation of one response unit exceeded a threshold. Because of the influence of the attention layer, the target stimulus had the strongest effect on the response units, which normally led to a correct response. Such a situation can be seen in Figure 2A, where the time course of activation of the units in the response layer is shown. A specific number of cycles after the first response, the spread of activation from the stimulus units to the response units was interrupted. This simulated the end of stimulus processing and is responsible for the fact that the response activations in Figure 2A decrease some time after the response.

Because of the noise, the activation of a wrong response could also exceed the threshold and, thereby, produce an error. This happened mainly when the noise led to a response before the attention layer could exert its influence on the input. Consequently, errors were typically faster than correct responses, something that has also been observed empirically (Luce, 1986). However, after such an error, the attentional set evolved on the given trial so that the correct response eventually exceeded the threshold. In this way, most errors were corrected (see Figure 2C). Only if the activation of the correct response failed to reach the threshold before stimulus processing was interrupted did an error remain uncorrected (see Figure 2B).

The network's ability to correct itself is the basis for error correction. We assumed that when a second response exceeds the threshold, this represents an *internal* correction response. Only if intended does this also lead to a corresponding *overt* correction response (i.e., to an ECR). If no overt error correction is intended, the ECR is suppressed even when an internal correction response has occurred. One could further assume that on a portion of trials, an internal correction response causes an unintended ECR. These spontaneous ECRs, however, were ignored in the present study (but see Fiehler et al., 2005). Generally, we did not consider the production and suppression of responses. Only the decisional part was modeled.

The RM Model

For modeling RM, we added a virtual RM-based error detector to the task module. That is, this mechanism was not implemented in the neural network but was realized by the method in which we computed the ESR performance (see Appendix A). Basically, we assumed that the RM-based error detector continuously monitors the response units and registers whenever a unit exceeds its threshold. When two different responses exceed the threshold in succession—or, in other words, when an internal correction response occurs—the error detector concludes that the first response was an error. If the system is instructed to signal its errors, an ESR_{RM} is initiated. The latency of the ESR_{RM} depends on the duration of the internal correction response and the duration of nondecisional processes related to the initiation and execution of the ESR_{RM}.

To explain why empirical ESRs occur less frequently and require more time than ECRs, we simply assumed that the nondecisional processes involved in the ESR_{RM} require more time and are more prone to failure than those involved in the ECR. We suggest that this is because of the fact that the ESR_{RM} additionally requires a switch to a different response system, which is not necessary for producing an ECR. This additional process not only requires time but also relies on central capacity and, as a consequence, is prone to distraction (see Rabbitt, 2002). Moreover, the system could simply "forget" to produce an ESR because the

[1] In the original model of Yeung et al. (2004) and Botvinick et al. (2001), the amount of attention that is directed to the target letters depends on the amount of response conflict in the previous trial. We adopted this mechanism, although this is not crucial for our model. However, the results we report do not depend on whether such a mechanism is implemented.

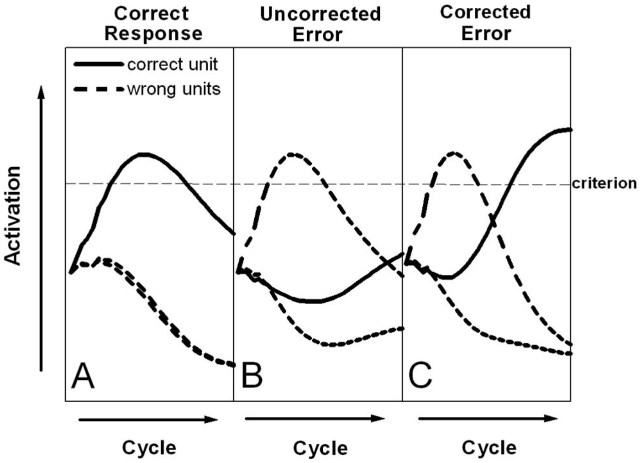

Figure 2. Idealized time course of response unit activation for trials with correct initial responses (A), trials with uncorrected errors (B), and trials with corrected errors (C). Solid lines indicate the activation of the response unit representing a correct response; dashed lines indicate the activation of the response units representing a wrong response. A response is selected whenever the activation of a unit exceeds the response criterion (dashed horizontal line).

respective goal is not active enough (in the sense of a goal neglect; e.g., De Jong, Berendsen, & Cools, 1999).

The CM Model

In contrast to the RM account, the model of Yeung et al. (2004) assumes that error detection relies on a separate mechanism based on CM. The core of this mechanism is a unit that registers the amount of conflict in the response layer, which is computed by a Hopfield energy measure, $E(n)$, at each cycle n (Hopfield, 1982):

$$E(n) = -\sum_{i}^{N}\sum_{j}^{N} act_i(n)act_j(n)w_{ij}.$$

Here, w_{ij} denotes the weight of the association between units i and j, with activations $act_j(n)$ and $act_i(n)$, respectively. The products between the weight and the activations are summed up for all N units in the response layer.[2] According to Botvinick et al. (2001), this measure captures the concept of a response conflict because it implies a high conflict when both responses are highly activated and a low conflict when only one response is highly activated or when no responses are activated. Yeung et al. (2004) used this formula to simulate the Ne/ERN. They calculated the mean response conflict in a time window beginning with the response. The Ne/ERN then equaled the difference in mean response conflict between correct trials and erroneous trials. They found that the simulated values were close to empirically observed Ne/ERN data.

Most important for the present objective, however, is the fact that response conflict was also used as basis for error detection. At each cycle, a counter was increased by the current amount of response conflict. When this *cumulated conflict* exceeded a certain threshold, an error was signaled. However, conflict accumulation did not start with the production of a response but only after a fixed delay. Yeung et al. (2004) found that, otherwise, a huge number of false alarms were produced. This occurred because, at the time a response exceeded the response criterion, a conflict was often present irrespective of whether the response was correct or not. This issue is of great importance for our objective and is discussed in more detail in the next section.

For implementation of our CM model, we equipped the task module with a CM-based error detector, as described above. It initiated the production of an $\mathrm{ESR}_{\mathrm{CM}}$ whenever an error was internally detected (i.e., whenever the cumulative postresponse conflict exceeded a threshold). The latency of the $\mathrm{ESR}_{\mathrm{CM}}$ consists of the time elapsed between an error and its detection as well as a nondecisional component that comprises the initiation and execution of the $\mathrm{ESR}_{\mathrm{CM}}$. The frequency of the $\mathrm{ESR}_{\mathrm{CM}}$ equals the frequency of trials on which an internal detection has occurred.

Taken together, our two models allow the calculation of two measures of error detection each: the ECR, which is identical in both models, the $\mathrm{ESR}_{\mathrm{RM}}$ and the $\mathrm{ESR}_{\mathrm{CM}}$. Whereas both the ECR and the $\mathrm{ESR}_{\mathrm{RM}}$ are based on internal correction responses, the $\mathrm{ESR}_{\mathrm{CM}}$ relies on a CM mechanism. In the next section, we consider how the models predict the ECR and ESR performance as a function of the response criterion.

Exploration of Model Behavior

The way we defined the RM model already implies that any variable that affects the internal correction response should affect ECR and $\mathrm{ESR}_{\mathrm{RM}}$ performance in a similar way. This, of course, also holds for the response criterion. Our aim in the present section is to examine whether the CM model would predict the same. In such a case, response criterion effects would not be used to distinguish between the models. A further goal of this section is to illustrate the mechanisms by which the response criterion affects the performance in our two models in general. This helps us later in interpreting the empirical results.

We simulated both models with a wide range of response criteria. In the following, we summarize the main results of these simulations. The details can be found in Appendix B. Indeed, the simulations revealed that the two accounts can be distinguished. It turned out that, as expected, the predictions mainly differed with respect to the latencies of ECRs and ESRs. In addition, we obtained some valuable insights into how the response criterion can affect error processing in general.

The simulations confirmed that reliable error detection can be achieved with both models. With reasonable parameters, the frequency of corrected and detected errors is rather high, whereas the rate of false alarms is sufficiently low. A first important question is how the response criterion affects the latencies of the initial response and of the internal correction response. The mechanisms underlying these responses are identical in both models. As ex-

[2] In case of only two response units, this measure equals the product of the activation of the two response units and the inhibitory weight multiplied by −2. We explored this formula for more than two units and found out that a meaningful response conflict results only with a slight modification: Only those pairs of units should be entered into the formula, for which each activation value is positive. Similarly, Yeung et al. (2004) defined the response conflict to be 0 if one of the two response units in their network had a negative activation.

pected, our simulations revealed a generally strong criterion effect on these variables. With a higher criterion, not only the initial response but also the internal correction response requires more time to exceed the criterion. Surprisingly, however, the criterion effect on the correction latency differed from that on the latency of the initial response.

The reason for this difference can be seen Figure 3, in which the averaged time course of response activation from trials with corrected errors is presented for two response criteria. Although the criterion effect on correction latency is mainly attributable to the fact that more time is required to reach a higher criterion, the criterion has two further effects on correction latency. First, the criterion effect is amplified by the fact that the response activation at the cycle where the error occurs also depends on the criterion. With a higher criterion, the activation of the response unit causing the error is higher and that corresponding to the correct response is lower. This implies that the subsequent correct response requires even more time to cause a correction.

Second, the criterion effect is slightly counteracted by the fact that a higher criterion implies a stronger activation built up for the correct response unit. Nevertheless, in our simulations, the criterion effect on ECR latency was generally stronger than that on the latency of the initial response. However, there might be conditions under which the criterion effect on ECR latency is even weaker than that on initial response latency. Such a case is shown in a later section. Altogether, we can conclude that ECR latency should show a criterion effect, which is not necessarily equal to that on the latency of the initial response.

With respect to the RM model, this implies that not only ECR latency but also $\mathrm{ESR_{RM}}$ latency should show such a criterion effect. This is a consequence of the fact that, according to the RM model, both measures depend on the duration of the internal correction response. The crucial question is whether the CM model makes a different prediction. Of course, the CM model would predict the same strong criterion effect on ECR latency, because ECR performance in the CM model is based on the same mechanism as in the RM model. Therefore, to distinguish between the models, it would be necessary to show that the prediction differs for the $\mathrm{ESR_{CM}}$ latency.

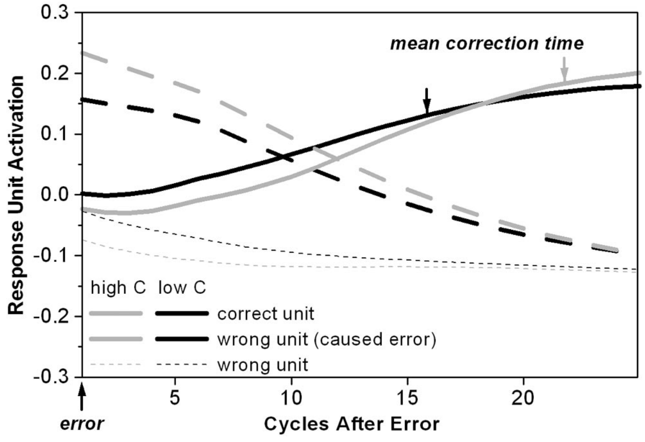

Figure 3. Mean activation of response units for trials with corrected errors, separately for conditions with a low and a high response criterion. Averaging was locked to the cycle on which the error response exceeded the criterion. Solid lines indicate activation of correct response units; dashed lines indicate activation of wrong response units (thick dashed lines indicate the wrong response unit that actually caused the error). Arrows mark the time of the error and the correction response. C = response criterion.

As expected, our simulations revealed that the response criterion also has an effect on the $\mathrm{ESR_{CM}}$ latency. The reason for this lies again in the response activation at the time the error response is selected. This is illustrated in Figure 4, where the averaged response conflict as well as the cumulated conflict is depicted for detected errors from conditions with a low and a high response criterion. Evidently, the response conflict at the time the error has occurred is reduced with a higher criterion. This is due to the fact that the activation difference between the correct and the wrong response is increased with a higher criterion, implying a reduced response conflict. If the measurement of cumulated conflict, on which error detection is based, would start at this time, one would obtain a rather strong criterion effect on error detection, implying a longer $\mathrm{ESR_{CM}}$ latency with a higher criterion.

However, the measurement of cumulated conflict does not start immediately after the error but, rather, after a specific delay, as illustrated in Figure 4. As discussed earlier, this is necessary to prevent the occurrence of false alarms. Since the criterion effect on response conflict decreases in the course of some cycles, it has nearly disappeared when cumulated conflict measurement starts. As a consequence, the criterion effect on error detection is only small. In Appendix B, we demonstrate that without such a delay, the criterion effect is stronger, but, at the same time, the false alarm rate is implausibly high. With a sufficiently long delay, the criterion effect on $\mathrm{ESR_{CM}}$ latency is far smaller than that on ECR latency. This suggests that the CM model makes a different prediction than the RM model. According to the CM model, $\mathrm{ESR_{CM}}$ latency should be less affected by the response criterion than is the ECR latency.

In addition, we found also criterion effects on the frequencies of ECRs and ESRs. It turned out that the frequency of correct ECRs (and, therefore, correct $\mathrm{ESR_{RM}}$s) slightly decreased with an increasing criterion. This is a side effect of the latency effect and results from the fact that the longer it takes until the internal correction response, the higher the probability that it fails to exceed the criterion before stimulus processing has terminated. However, the CM model would predict the same results, at least under specific conditions (see Appendix B). As a consequence, if we found such an effect empirically, this would not distinguish between the models.

Taken together, our simulations confirmed that varying the response criterion is useful for testing between the two accounts. If the RM account is valid, ECR and ESR performance should show generally similar criterion effects for the latencies as well as for the frequencies. In contrast, the CM account would predict different criterion effects for both responses, at least for the latencies. According to this account, the criterion effect should be much stronger for the ECR latency than for the ESR latency.

Experiment

To test the derived predictions for the CM and RM accounts, we conducted an experiment in which an Eriksen-flanker task corresponding to our model was used. Participants had to classify a

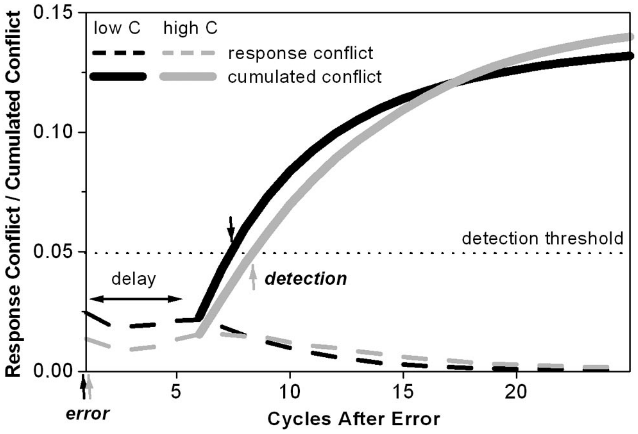

Figure 4. Mean response conflict and cumulated conflict for trials with errors detected by the conflict monitoring model, separately for conditions with a low and a high response criterion. Averaging was locked to the cycle on which the error response exceeded the criterion. The delay represents the time by which conflict accumulation was deferred after the initial response (parameter *D* in the model). Arrows mark the time of the detection response. C = response criterion.

target letter by pressing one of three response buttons with fingers on one hand. The target was flanked either by letters of a different category (incongruent condition) or by neutral symbols (neutral condition). In half of the blocks, the participants were instructed to give an ESR by pressing a neutral key with the hand not used for the main task (ESR condition), whereas in the other half, they were instructed to correct their errors (ECR condition).[3]

The response criterion was varied by means of a deadline procedure in which the participants were instructed to give their first response before an acoustical signal sounded. The interval from stimulus to signal onset (i.e., the deadline) varied among three levels across blocks. In this way, the participants could anticipate the deadline and adjust their response criterion in an optimal way. An alternative procedure would have been to instruct the participants to emphasize either speed or accuracy. However, such a method might have induced further strategic differences. Yeung et al. (2004), for instance, speculated that emphasizing accuracy versus speed could also lead to different degrees of attention.

Method

Participants. Ten participants (3 female, 7 male) between 19 and 30 years of age ($M = 25.2$) with normal or corrected-to-normal vision participated in the study. All were right-handed. Participants were recruited at the Universität Konstanz, Konstanz, Germany, and were paid €5 (U.S. $6) per hour.

Apparatus. The stimuli were presented on a 21-in. (53.34-cm) color monitor. An IBM-compatible PC controlled stimulus presentation and response registration.

Stimuli. Stimulus arrays were composed of a target letter and two identical distractor letters, which were on the left and on the right of the target. The letters *K*, *V*, *M*, *P*, *R*, and *W* and the neutral symbols %, &, #, and § were taken from an Arial font and resized on a visual angle of 1.67° height and 1.51° width at a viewing distance of 127 cm. The whole array subtended a visual angle of 5.10° width. Two letters were assigned to one response each. Each letter was used as a target letter and was combined with a distractor letter either from the set of the four letters that required a different response (incongruent stimulus) or from the set of neutral symbols (neutral stimulus). In this way, 48 stimuli were constructed.

Procedure. Participants were told to respond to the identity of the target and to ignore the flanker letters. Responses were given with the fingers of the right hand. Depending on the letter, a keypress with the index finger was required if the target was either the letter *K* or *V*. A keypress with the middle finger was required if the target letter was either the letter *M* or *P*. Otherwise, a keypress with the ring finger was required.

Each trial started with a stimulus array presented for 150 ms, followed by a blank screen. After a specific interval, an acoustical deadline signa1 (800 Hz) sounded for 150 ms. Participants were instructed to respond faster than this signal. In half of the blocks, participants were also instructed to correct their errors by pressing the correct key immediately after they had detected the error (ECRs). In the other half of blocks, they had to signal errors by pressing the space bar of a standard keyboard with their left hand immediately after each error (ESRs). Following an interval of 1,500 ms after the first response, a new trial started. If further responses (ECRs, ESRs) occurred within this interval, a new interval of 1,500 ms was started. No feedback on the accuracy of the response was provided. However, on some trials a speed feedback was given. Whenever the response time exceeded the deadline on five consecutive trials, the German word *schneller* (faster) was presented for 200 ms on the screen 200 ms after the response.

Each block consisted of 96 trials, 2 for each possible stimulus. Half of the stimuli were neutral, and the other half were incongruent. Participants worked through 24 test blocks distributed across two test sessions for a total of 2,304 trials. The ESR and ECR instructions alternated between blocks. The type of the first instruction was counterbalanced across participants. Furthermore, there were three deadline conditions (low, intermediate, high), which were constant within each block but varied across blocks. The order of the deadline condition was randomized, and 4 blocks of each deadline occurred in each session.

Each session started with 3 practice blocks, followed by 12 test blocks. In a preliminary practice session, 12 practice blocks were performed. In the first 4 blocks of this practice session, no deadline was applied. Rather, these blocks served to determine the three deadlines in subsequent blocks. For each participant, the intermediate deadline was individually set to the median response time in the 4th practice block. The low and high deadlines were obtained by subtracting 50 ms from and adding 50 ms to the intermediate

[3] Alternatively, we could also have used spontaneous error corrections as a measure of ECR performance (Fiehler et al., 2005). However, there is evidence that error corrections are actively suppressed if they are not instructed (Rabbitt & Rodgers, 1977; Steinhauser & Hübner, 2006). Accordingly, not all internal corrections would have led to an ECR. It is even possible that the number of inhibited ECRs depends on the deadline. Such inhibitory mechanisms should not be involved in error signaling, because it is reasonable to assume that the production of an ESR is not automatic. Thus, the use of noninstructed ECRs would have confounded the two measures with respect to the presence of inhibitory mechanisms.

deadline, respectively. These deadlines were used throughout the entire experiment.

Results

To control for outliers, trials were excluded whose first response time was 2 standard deviations above or below the mean (<1%). The remaining trials were classified with respect to whether the first response was correct or wrong and whether it was followed by an ECR or an ESR. Trials that included more than two responses were excluded with one exception: Errors that were followed by an ECR as well as an ESR were assigned to a separate category. As an overview, Table 1 reports the relative frequencies of ECRs and ESRs within trials with correct and erroneous responses for our two main conditions.

The table reveals a high number of spontaneous error corrections in the ESR condition (35%). Because we do not know how ESR latency is affected by a preceding ECR, the following analyses included only trials from the ESR condition in which no ECR was involved. However, separate analyses revealed that trials with spontaneous ECRs showed a rather similar pattern, although the low absolute number of trials with both an ECR and an ESR made a stable estimation of latencies difficult.

Below, we report analyses of those dependent variables that were used to test the predictions of the model. We start by reporting analyses for each dependent variable separately. In a final section, we compare ECR and ESR performance. Figure 5 depicts each dependent variable as a function of deadline level.

Table 1
Frequencies and Latencies for all Trial Types Observed in the Experiment

	Initial response correct		Initial response wrong	
Condition and response	Freq. (%)	RT of consecutive responses (ms)	Freq. (%)	RT of consecutive responses (ms)
Condition ECR				
No ECR	99.0	526	13.9	489
ECR	1.0	—/—	85.2	492/412
Wrong ECR			1.0	—/—
Condition ESR				
No ECR				
No ESR	98.3	532	12.6	—
ESR	0.9	—/—	50.9	510/557
ECR				
No ESR	0.5	—/—	13.5	530/168
ESR	0.3	—/—/—	21.5	502/317/844
Wrong ECR				
No ESR			0.3	—/—
ESR			1.3	—/—/—

Note. Frequencies were computed relative to all trials in which the initial response was either correct or an error within conditions in which ECRs (error correction responses) or ESRs (error signaling responses) were required. For the latencies, the first value represents the latency of the initial response; the second value represents the latency of the ECR or the ESR (when no ECR occurred), computed as the difference between the initial response and the ECR/ESR; and the third value represents the latency of an ESR, computed as the difference between the initial response and the ESR. ESRs followed by ECRs are not considered because they were virtually never observed. Dashes indicate that no latency could be calculated because of too few trials or because some participants had empty cells. RT = response time; Freq. = frequency.

Initial responses. To check whether our manipulation of response criterion was successful, we analyzed the response times of correct responses and the overall error rate. Although we focused on the effect of the deadline, we also included the variable stimulus congruency in these analyses. In examining the influence of the flanker letter, we wanted to test whether our participants applied strategies other than a mere criterion shift to adapt to the different deadline levels. For instance, if a long deadline implies that more attention is directed to the target than with a short deadline, we should observe a decreased congruency effect in this condition.

To calculate the mean latency of correct responses, we averaged the latencies of initial responses from trials in which the initial response was correct. The data were entered into a three-way ANOVA with repeated measurement on the variables block type (ECR condition, ESR condition), deadline (1, 2, 3), and congruency (neutral, incongruent). The analysis revealed significant main effects of all variables. Mean response time was increased in the ESR condition (521 ms) relative to the ECR condition (513 ms), $F(1, 9) = 21.3$, $p < .01$. It was increased with incongruent stimuli (523 ms) relative to neutral stimuli (510 ms), $F(1, 9) = 60.9$, $p < .001$. Finally, it increased linearly with an increasing deadline level (1: 487 ms; 2: 516 ms; 3: 547 ms), $F(2, 18) = 137.7$, $p < .001$. No significant interactions were obtained.

The error rate denotes the relative frequency of erroneous initial responses. The data were subjected to the same type of analysis outlined above. The analysis indicated significant main effects of deadline, $F(2, 18) = 31.5$, $p < .001$, and congruency, $F(1, 9) = 88.1$, $p < .001$, representing the fact that the error rate decreased linearly with an increasing deadline level (1: 27.6%; 2: 19.0%; 3: 13.6%) and was higher on incongruent trials (23.3%) than on neutral trials (16.8%). However, these effects were qualified by a significant Block Type × Deadline × Congruency interaction, $F(2, 18) = 7.28$, $p < .01$. This can be attributed to the fact that the performance difference between neutral and incongruent stimuli fluctuated in a nonsystematic manner between the different deadline conditions of the ESR blocks (1: 4.7%; 2: 9.3%; 3: 5.7%) and the ECR blocks (1: 8.5%; 2: 5.4%; 3: 5.3%).

ECR and ESR performance. ECRs and ESRs were analyzed in a similar way. For each measure, three dependent variables were calculated and entered into a one-way ANOVA with repeated measurement on the variable deadline (1, 2, 3). The latency was calculated as the time elapsed between the erroneous response and the respective detection response, including only trials in which errors were successfully corrected or signaled. The hit rate was calculated as the relative frequency of successfully corrected or signaled errors relative to the rate of all trials where the initial response was an error. Finally, the false alarm rate was calculated as the relative frequency of erroneously corrected or signaled correct responses relative to the rate of all trials with a correct initial response.

The analysis of the ECR latency revealed a significant effect of deadline, $F(2, 18) = 14.6$, $p < .001$. The mean correction time increased with an increasing deadline level (1: 371 ms; 2: 415 ms; 3: 479 ms). The ECR hit rate showed a marginally significant effect of deadline, $F(2, 18) = 3.38$, $p = .057$. The hit rate was similar on the first two deadline levels but decreased on the third

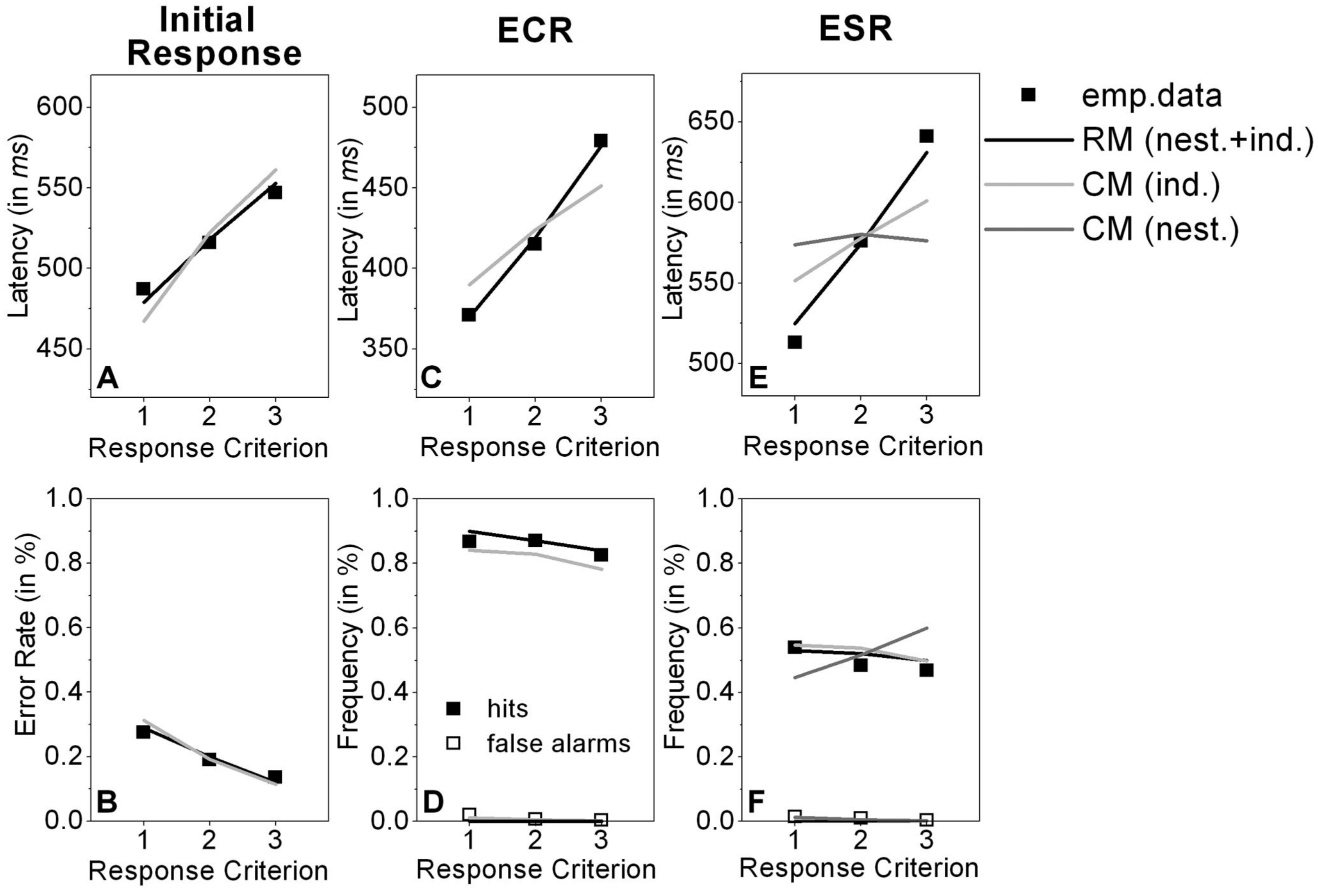

Figure 5. Effect of response criterion on the latency of the initial response on correct trials (A), the error rate (B), ECR latency (C), ECR hits and false alarms (D), ESR latency (E), and ESR hits and false alarms (F) in the empirical data, the RM model fit, and the CM model fit. Whereas for the CM model, the independent (ind.) and nested (nest.) fits are shown separately, both fit types were identical for the RM model (for the nested CM model fit, only ESR data are shown, because initial response data and ECR data correspond to those of the RM fit). ECR = error correction response; ESR = error signaling response; RM = response monitoring; CM = conflict monitoring.

level (1: 86.8%; 2: 87.1%; 3: 82.6%). Finally, the analysis of the ECR false alarm rate revealed a nonsignificant trend toward an increased false alarm rate on the lowest deadline level (1: 2.1%; 2: 0.8%; 3: 0.4%).

The analysis of the ESR latency showed that deadline had a significant effect on this variable, $F(2, 18) = 10.2$, $p < .001$. The detection time increased linearly with an increasing deadline level (1: 513 ms; 2: 576 ms; 3: 641 ms). The analysis of the ESR hit rate showed no significant effect of deadline. However, there was a nonsignificant trend toward a reduction of the detection hit rate at higher deadline levels (1: 53.9%; 2: 48.2%; 3: 46.9%). Finally, the analysis of the ESR false alarm rate revealed a significant effect of deadline, $F(2, 18) = 3.78$, $p < .05$. False alarms were more frequent at lower deadline levels (1: 1.5%; 2: 0.9%; 3: 0.4%).

Comparison of ECR and ESR. To compare ECRs and ESRs, we entered both into the same analyses. We computed two-way ANOVAs with repeated measurement on the variables deadline (1, 2, 3) and measure (ESR, ECR). Only values involving the variable measure are reported. For the latencies, the main effects of deadline, $F(2, 18) = 26.2$, $p < .001$, and measure, $F(1, 9) = 76.2$, $p < .001$, were significant. ESR latency (577 ms) was increased relative to ECR latency (422 ms). The Deadline × Measure interaction was not significant. For the hit rates, only the main effect of measure reached significance, $F(1, 9) = 118.2$, $p < .001$. The hit rate for ESRs (49.7%) was reduced compared with that for ECRs (85.5%). Again, no significant interaction was detected. For the false alarm rates, no significant effect was obtained.

Discussion

In the present experiment, the response criterion was manipulated by varying a response deadline between blocks. We hoped that the participants would use an individual criterion for each deadline. The analyses of the latencies and error rates for the initial responses confirmed that the manipulation was successful. Response times of correct responses increased linearly with the deadline, whereas the error rates decreased. Moreover, there is no evidence that our participants adopted different attentional strategies for the different deadline levels. This can be concluded because the distractors had a similar effect for the different deadline levels, at least for the response times. Taken together, it seems that the deadline effects in our paradigm were attributable to a shift in the response criterion.

ECR and ESR performance was measured in two blocked conditions in which participants were either instructed to correct their errors or to signal their errors. A preliminary analysis revealed that a substantial number of errors were spontaneously corrected in the ESR condition, although this was not instructed. Possibly, the tendency to spontaneously correct errors is facilitated when an ESR is required (e.g., Steinhauser & Hübner, 2006). Interestingly, more than one third of these spontaneously corrected errors did not lead to an ESR, although this was instructed. However, the correction latency for these unsignaled errors was very short (168 ms). Accordingly, one could hypothesize that the participants were unsure whether they should signal an error when this error was almost immediately corrected.

Most important for the present objective are the deadline effects on ECR and ESR performance. On the basis of our simulations, we derived different predictions for the CM account and the RM account. According to the RM account, we should have observed similar effects of deadline on ECR and ESR performance. In contrast, the CM model predicts different effects, at least for the latencies. According to this model, the deadline effect on the ESR latency should be weaker than that on the ECR latency. Our data clearly support the RM account. ESR and ECR performance was similar, not only for the latencies but also for the hit rates and false alarm rates. We observed only a difference with respect to the absolute latency and hit rate. ESRs required more time and were slower than ECRs, which replicates the result of earlier studies (Rabbitt, 2002). This, however, is also consistent with an RM account, if we assume that this difference is attributable to an increased failure probability and duration of selecting the signaling response.

There are other interesting results in our data. First, the deadline effect on the ECR latencies was stronger than that on the latency of the initial response. Second, the ECR and ESR hit rates slightly decreased with an increasing deadline. Finally, the false alarm rates were rather low. Although these findings cannot distinguish between the CM and RM model, they are consistent with the results of our simulations.

The fact that our experiment confirmed the predictions of the RM model in nearly every detail shows the model's high validity. In contrast, the CM model did not predict the similar criterion effects on ECR and ESR latency. However, we varied only a few parameters in our simulations (see Appendix B). Most of them were fixed to values used by Yeung et al. (2004) to simulate Ne/ERN data. This raises the question of whether other parameter values would have also allowed the CM model to predict our empirical results.

To see whether this is the case, we fitted the models to our data by means of an exhaustive parameter search. Usually, this procedure is difficult for connectionist models, because there are two problems. First, model performance is strongly influenced by noise, which reduces the efficiency of search algorithms. Second, the number of parameters is often very large in connectionist models, requiring a high computational effort to search parameter space. Fortunately, Bogacz and Cohen (2004) introduced a search procedure that deals with the problems inherent in neural networks. To keep the computational effort low, we optimized only those parameters that we considered relevant for the present purpose.

Model Fit

Each model was fit to the data of eight empirical variables: the latencies, hit rates, and false alarm rates of ECRs and ESRs; the latencies of correct responses; and the overall error rate. Table 2 gives an overview of the parameters that were fit for each model. Ten parameters were the same for the RM model and the CM model: the three response criteria (C_{low}, C_{med}, C_{high}); the time after which stimulus processing was stopped following the first response (d_{stop}); the time constants corresponding to the duration of nondecisional processes of the initial response, ECR and ESR (T_{ND1}, T_{ND2}, T_{ND3}); the time per cycle (T_{cycle}); and two scaling factors weighting the strength of inhibitory (sc_{i}) and excitatory (sc_{e}) connection weights. For the RM model, the additional pa-

Table 2
Free Parameter Values for the Best Fits of Our Models and for Yeung et al.'s (2004) Original Model

Model parameter	Original	RM best fit (nest. + ind.)	CM best fit (nest.)	CM best fit (ind.)
Response criterion low (C_{low})	—	0.176	0.176	0.150
Response criterion intermediate (C_{med})	0.18	0.204	0.204	0.200
Response criterion high (C_{high})	—	0.236	0.236	0.235
Inhibitory scaling (sc_{i})	0.08	0.125	0.125	0.256
Excitatory scaling (sc_{e})	0.12	0.140	0.140	0.121
Time at which stimulus processing is interrupted after first response (d_{stop})	6	2.02	2.02	8.16
Time per cycle (T_{cycle})	16	27.36	27.36	18.84
RT1 time constant (T_{ND1})	200	208.9	208.9	244.0
ECR time constant (T_{ND2})	—	70.7	70.7	67.2
ESR time constant (T_{ND3})	—	227.5	210.6	295.3
ESR failure rate (P[ESR fails])	—	0.404	—	—
Detection delay (D)	6	—	10	4
Detection threshold (K)	Variable	—	0.132	0.028

Note. Dashes indicate parameters that are not available for the particular model. RM = response monitoring; CM = conflict monitoring; nest. = nested; ind. = independent; RT = response time; ECR = error correction response; ESR = error signaling response.

rameter P(ESR fails) was used, which specifies how frequently the selection of the ESR fails, despite the occurrence of an internal correction response. For the CM model, the detection delay D and the detection threshold K were additionally optimized. All other parameters (e.g., connection weights) were fixed to the values used by Yeung et al. (2004). Please note that the response criterion C was the only parameter that was allowed to vary between the criterion conditions. All other parameters were held constant across these conditions.

We applied the search procedure introduced by Bogacz and Cohen (2004), which they explicitly developed for neural networks like the present one. The algorithm proceeds in three phases: an initial parameter search, an optimization phase, and a tuning phase. Each of these phases consists of a fixed number of iterations. We set the number of iterations for each phase to 400, 200, and 100. In each iteration, the respective model was calculated 5,000 times for each stimulus type (incongruent, neutral) within each of the three response criterion conditions. Again, the results were averaged across the two stimulus types. The whole algorithm was applied 10 times for each stage of fitting.

To estimate the quality of the fits, we calculated as goodness-of-fit statistics the mean squared errors. Because our data consisted of two types of measures (latencies, frequencies) that are different in magnitude, we corrected each difference between empirical and model value by multiplying it with a correction factor. This correction factor was 1 for latencies and 500 for frequencies. In this way, an error of 1% in a frequency measure corresponded to an error of 5 ms in a latency measure. Accordingly, the mean square error was calculated as

$$MSE = \sum_{i} [(emp_i - sim_i)n_i]^2,$$

where emp_i is the empirical data point i, sim_i is the simulated data point i, and n_i is the correction factor of the pair of data. From the 10 applications of the fitting algorithm to each model (and stage of fitting), we chose the outcome with the lowest mean square error as the best fit.

The models were fitted by applying two strategies. The first strategy was to fit each model independently to the data (*independent fits*). This method implied that the whole set of parameters was separately fitted to all empirical variables for each model. The second strategy was to fit parts of both models simultaneously in a nested manner (*nested fits*), a process that proceeded in two stages. In the first stage, we used the parameters shared by both models and fitted the latency and frequency measures related to the initial response as well as those related to ECR performance. In the second stage, we separately estimated the additional parameters of the RM account to obtain the ESR_{RM} (i.e., T_{ND3} and P[ESR fails]) and the additional parameters of the CM account to obtain the ESR_{CM} (i.e., T_{ND3}, K, and D).

We used the nested fit method to obtain comparable parameter values for each model. Moreover, in this way, the two model fits differ exclusively with respect to the ESR performance, which is crucial for our reasoning. However, this method is disadvantageous for the CM model. Optimizing all parameters for the first response and ECR performance generally constrains the model's ability to fit the ESR. These constraints, however, are less strong for the ESR_{RM}. Since the criterion effects on the empirical ECR performance are similar to those on the empirical ESR performance, a good fit of the RM model to the ECR performance always implies a good fit to the ESR performance. For the CM model, in contrast, it is possible that the better the model fits the criterion effect on ECR performance, the worse it fits the criterion effect on ESR performance. Therefore, we also had to use the independent-fit strategy to guarantee equal opportunities for both models to fit the data.

Results and Discussion

Figure 5 shows the data of each fit together with the empirical data. The parameter values obtained for the best fit of each model and fitting strategy are presented in Table 2.

We first calculated the independent fits. The best fit of the RM model produced a mean squared error of 4,256, which is superior to the best fit of the CM model (MSE = 13,574). However, there is also an important qualitative difference. Visual inspection of Figure 5E reveals that the CM model is not able to model the criterion effect on ESR performance. As expected, the slope of the predicted criterion effect for the ESR latency is too flat, and most important, the criterion effect on ESR latency is smaller than that on ECR latency. Interestingly, the search algorithm's attempt to optimize the prediction of ESR performance worsened its prediction of the ECR performance (relative to that of the RM model). Because this procedure did not assign different priorities to the fit of ECRs and ESRs, the search algorithm yielded a better overall result by improving the fit to the ESR data at the cost of an impaired fit to the ECR data. This can be illustrated by considering how ESRs and ECRs contribute to the overall mean square error of the latency measures. Only 64% of the latencies' mean square error in the CM model is attributable to ESRs. However, 25% results from ECRs. This shows that a large portion of the fit error in the CM model is due to bad fit of ECR latencies.

In a further step, we calculated the nested fits. Fitting the shared parameters of both models to the data for the initial response and ECR resulted in the same parameters as those of the independent RM model fit. As a consequence, the independent and nested fits are identical for the RM model. On the basis of these parameters, we fitted the remaining parameters for the CM model. As expected, the fit of the CM model to the ESR data was now even worse (see Figures 5E and 5F). Whereas the predicted criterion effect on ESR latency was clearly too small, the predicted criterion effect on ESR hit rate was even reversed. The observed difference between the fit of both models is reflected in the different mean square error values (which we now calculated for ESR performance only). The mean square error for the RM model was much smaller (934) than that for the nested CM model (14,660). Moreover, the latter mean square error is also clearly higher than that of the ESR data from the independent CM model (4,131).

Taken together, the fits of the models confirm our conclusions from the initial simulations and from our experiment. The independent fits of both models show that the RM model but not the CM model properly accounts for the effects of response criterion on all aspects of performance. However, the nested fits show that it is indeed ESR performance that is crucial for distinguishing between these models. If we force both models to adopt parameters that optimally fit initial response and ECR data, the RM model can also account for the ESR data, whereas the CM model has severe problems in achieving this.

General Discussion

The present study addressed the question of whether error detection, as measured by the ESR, is based on RM or CM. Whereas RM relates error detection to the detection of an internal correction response, CM assumes that this is achieved by detecting a response conflict. We have argued that the two accounts can be distinguished by comparing the effects of the response criterion on error signaling and error correction performance. To derive differentiated predictions for the two accounts, we implemented them as extensions of the neural network model developed by Yeung et al. (2004). According to the RM account, both ESRs and ECRs rely on the same internal correction response. Consequently, this account predicted that both responses should generally show the same criterion effect. In contrast, the CM account predicted a smaller criterion effect for ESR latency than for ECR latency, because ESRs rely on conflict monitoring, which turned out to be less sensitive to the response criterion than were ECRs.

To test these predictions, we conducted an experiment with a three-alternative forced-choice version of the Eriksen-flanker paradigm. Three response alternatives were essential for our objective, because a true error correction requires that more than two responses be used. We manipulated the response criterion by varying a response deadline. We assumed that the participants would adopt a separate response criterion for each possible deadline. This was plausible, because the deadline was varied between blocks, which allowed the participants to adjust the criterion in advance.

Our results obtained with this procedure confirmed the predictions of the RM model. The criterion effects on the latencies and on the error rates were the same for ESRs and ECRs. Moreover, a fit of the models to our data revealed that the CM model was not capable of accounting for our empirical deadline effects, whereas the RM produced an excellent fit. As one would have expected from our exploratory simulations, the CM account had great problems predicting the strong deadline effects on the ESR performance. It could be argued that we used a relatively large number of parameters (11 and 12 for the RM and CM models, respectively) to account for our 24 data points. However, it was not our goal to fit the data with a minimum number of parameters. Rather, we wanted to show that the RM model can account for the data and that the CM model cannot, even with many free parameters.

The RM model could also account for other aspects of our data. For instance, the criterion effect on ECR and ESR latency turned out to be stronger than that on the latency of the initial response. According to the RM model, a higher criterion implies that the correct response is suppressed more strongly when an error has occurred, which prolongs the generation of an internal correction response. This amplifies the criterion effect on the internal correction response relative to that on the initial response.

A second observation was that the frequency of correct ECRs and ESRs decreased with an increasing response criterion. According to the RM model, this was due to the increased latency of the internal correction response. The more time the correction response took, the higher the probability that stimulus processing terminated before the correction response had exceeded the criterion. Thus, application of the computational model not only allowed us to distinguish between two competing models, it also provided explanations of unexpected aspects of ECR and ESR performance. This demonstrates the strength of this approach.

Our results support the RM account as a mechanism underlying the behavioral measures of error detection. However, there are some critical issues that we have not discussed so far. First, our approach was restricted to specific implementations of the RM and CM accounts. Other models or other implementations of the CM theory were not considered. Second, our RM model focuses mainly on effects of response criterion. It is less elaborated with respect to other aspects of our data. Finally, our consideration of the CM theory focused exclusively on its capacity to explain error detection. However, this theory was initially developed to account for other phenomena like the Ne/ERN. In the sections below, we address these three topics.

Alternative Models?

One might ask whether there are, apart from RM and CM, alternative accounts that could explain our data? A possible candidate is a neural network developed by Holroyd, Yeung, Coles, and Cohen (2005) for modeling error detection in the Eriksen-flanker task, which implements the reinforcement learning theory of the Ne/ERN (Holroyd & Coles, 2002). According to this theory, the Ne/ERN indicates a negative reinforcement signal, which results whenever an event occurs that is at odds with an internally generated expectation. This happens, for instance, when an error occurs although the correct response is known.

Because the model of Holroyd et al. (2005) contains a mechanism for error detection, it is in principle capable of simulating the ESR. The model consists of a task module, which resembles that of the present model, and a monitoring module. By continuously evaluating the state of the task module, the monitoring module detects errors. This, however, is achieved without relying on response conflicts or internal corrections. Rather, the monitoring module signals an error whenever a stimulus and response that do not correspond to the instructed mapping are activated concurrently. In other words, errors are detected because the monitoring module already knows the correct response for a given stimulus. In this way, however, the detection of an error should not depend on the response criterion at all, because this plays a role only in the task module. Therefore, if an ESR were to rely on this type of error detection, one would not predict a criterion effect on ESR performance.

One could also ask whether Yeung et al. (2004)'s CM account could be modified in such a way that it would be consistent with our data. As already discussed, there is no direct way for the response criterion to affect the latency of CM-based error detection, because the response criterion does not affect the time course of conflict directly. There is only an indirect way. At the cycle at which the erroneous response exceeds the criterion, conflict depends on the response criterion, because the selected response is less activated with a low criterion than with a high criterion. In the present model, this can strongly affect ESR latency only when the accumulation of conflict starts immediately. However, as we show in Appendix B, with such a 0 delay of conflict accumulation, false alarms are very frequent, because the conflict at the time of the initial response can also be high for a correct response. Thus, the question is whether it is possible to construct a model in which the response criterion strongly influences ESR latency without suffering from an increased false alarm rate.

A quite different possibility for how the CM model could be modified would be to introduce additional assumptions regarding which parameters vary between the different conditions. One could simply assume that manipulating the response criterion implies also that the parameters of error detection change. For instance, a higher response criterion could be accompanied by a higher detection criterion (*K*) and a higher delay (*D*). In this way, the system could optimize the detection process by maximizing hits and minimizing false alarms. As a side effect, ESR latency would increase with an increasing response criterion.

We cannot exclude the possibility that a modified CM model exists that can account for the present data. However, it would probably require many additional assumptions regarding, for instance, the relation between the detection criterion and the response criterion. Moreover, any modification of the model architecture should maintain its capacity to account for the other phenomena such as the ECR performance or the Ne/ERN. Finally, even if the CM model can be modified in such a way as to enable it to explain the observed strong criterion effect on ESR latency, it would not necessarily predict that the criterion effect on ESR performance is the same as that on ECR performance. This is because even a modified CM model would have to assume that ESRs and ECRs are based on different mechanisms.

In contrast, an RM-based error detector can account for our results very robustly. The RM account should predict the present results for any implementation and architecture (e.g., a diffusion model), because each version would predict that ECRs and ESRs both rely on internal error correction and that the latency of this process depends on the response criterion. Thus, with respect to Occam's razor, the RM account is generally the better model in the context of our data.

Mechanisms Underlying ESRs and ECRs

The main goal of the present study was not to develop an elaborated model of RM. Rather, the present RM model was meant to account for criterion effects on ESR and ECR performance, which we identified as crucial for distinguishing between RM and CM accounts of error detection and which the CM model failed to account for. Unfortunately, the RM model is less specific regarding some other aspects of ECR and ESR performance. We had to make several strong assumptions to explain the differences between ECRs and ESRs. In the following, we summarize these assumptions and discuss their plausibility.

The RM model implies that although ECRs and ESRs are based on internal correction responses, they differ with respect to the processes that occur after an internal correction response is selected. More specifically, it is assumed that when ECRs are instructed, the correction response can directly be transformed into an overt response. In contrast, when ESRs are instructed, the internal and external responses have to be compared, and the signaling response has to be selected and initiated. We concluded that ESRs require more time than ECRs because they require more additional processes. Most important, the nature of the additional processes involved in ESRs (decision about and selection of a response) implies that they also require more capacity than do those involved in ECRs. This could explain why specifically ESRs are more prone to interference than are ECRs (Rabbitt, 2002) and, therefore, why ESR production sometimes fails.

Nevertheless, it is a weakness of the present RM model that it has to include a free parameter, *P*(ESR fails), to explain the different hit rates of ESRs and ECRs. Without this parameter, the model would hardly provide a good fit to the hit rate of error signaling, because the RM model has no other way to account for different rates of ESRs and ECRs. Moreover, to account for the data, we also had to assume that *P*(ESR fails) is independent of response criterion. Such an assumption is plausible if we assume that the failure to produce an ESR is related to processes outside of response selection for the main task (e.g., the selection and initiation of the signaling response) and, therefore, should be independent of response criterion of the main task. Unfortunately, however, this explanation implies that an important aspect of ESR performance is not captured by the simulation part of our model.

At first glance, the CM model is better suited to explain absolute differences between ESR and ECR performance, because it attributes ECRs and ESRs to different processes. However, a closer look at our simulation results reveals that this holds only partially. Indeed, CM-based error detection can also account for the observed data only with further assumptions. A CM-based error detector is faster than an RM-based error detector (see Appendix B), because the posterror conflict is highest before the internal correction occurs. Moreover, provided a liberal detection criterion, the hit rate of CM-based error detection is generally higher than that of RM-based error detection, because correction implies a response conflict (the activation curves of the responses always cross before correction occurs; see Figures 2 and 3), whereas a conflict can occur without a subsequent correction. Therefore, one would expect that ESRs based on CM would be faster and more frequent than ECRs. To explain the opposite pattern, the CM account would have to make additional assumptions, similar to the RM account.

A further strong assumption of the present RM model concerns the distinction between internal and overt error correction. We assumed that internal error correction is independent of whether participants are instructed to produce an ECR or an ESR. This, in turn, implies that ECRs can be suppressed without affecting the internal correction process. Basically, such an assumption requires that one distinguish between the selection of a response (which is necessary to determine the internal correction response) and the initiation of this response. Indeed, such an assumption can be derived from stage models of choice tasks that distinguish between a response selection stage and a response production stage (e.g., Pashler, 1984).

Nevertheless, some results suggest that our assumptions might be too simple. For instance, it is frequently observed that ECRs cannot be suppressed completely. In our experiment, participants often corrected their errors, even though this was not instructed. Recently, Fiehler et al. (2005) distinguished between automatic and intentional corrections to account for this phenomenon. To explain this within our RM model, we have to make further assumptions. For instance, the instruction not to correct errors could imply that the response channels are blocked after the first response is produced. Moreover, this suppression could require some time. Therefore, an automatic correction could occur whenever the correction response is selected before the blocking of response channels is finished. This would also explain why automatic corrections are faster than intended corrections.

Crucial, however, is the assumption that the blocking of ECRs takes place on the level of response production, whereas the present model simulated response selection. When a response exceeds the threshold in our model, this does not imply that this response is produced but, rather, that the response is selected. Because of this, suppressing an ECR does not affect the course of response activation during response selection. This contradicts a recent idea of Ullsperger and von Cramon (2006). These authors observed that when participants were instructed not to correct (but to signal) their errors, they responded more slowly but more accurately. Because of this, Ullsperger and von Cramon proposed that ECRs are prevented by increasing the response criterion. However, we did not observe such a speed–accuracy trade-off in our data (see also Steinhauser & Hübner, 2006). This suggests that a change in response criterion might support the suppression of ECRs under some conditions (e.g., by reducing the risk of automatic corrections), but it is not the only possible mechanism.

Implications for the CM Framework

A final note concerns the idea of conflict monitoring in general. Although the CM account of error detection was tested and rejected in the present study, this is not a rejection of the CM framework as a whole. For instance, implementing an RM account of error detection in the model of Yeung et al. (2004) does not change the model's capacity to simulate the Ne/ERN by means of response conflict (which initially motivated the CM model). Moreover, CM was not originally introduced to explain error detection alone. Rather, it has been suggested that CM supports the adaptation of control states. For instance, Botvinick et al. (2001) showed in a number of simulations how CM can be used for the flexible adjustment of attentional set.

However, our study demonstrates that each possible function of CM has to be examined independently. The fact that response conflict can account for an error-related phenomenon such as the Ne/ERN does not imply that response conflict is also involved in the detection of errors. In the same way, our conclusion that conscious error detection is related to RM does not imply that the Ne/ERN is also related to this process. Indeed, recent evidence suggests that the Ne/ERN is independent of whether an error is consciously detected or not (Endrass, Franke, & Kathmann, 2005; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). Therefore, our results do not necessarily support RM-related accounts of the Ne/ERN (e.g., Falkenstein et al., 1990; Gehring et al., 1993).

One possibility is that RM and CM coexist as two evaluative mechanisms with different characteristics and different functions. RM is relatively slow but very reliable for detection of an error. In contrast, CM delivers a faster evaluation of current processing demands. Accordingly, CM could serve as an early alerting mechanism. However, CM is possibly not reliable enough for a conscious error detection. As discussed earlier, a CM-based error detector is rather susceptible to false alarms. It required an additional mechanism (the delay in conflict accumulation) to counteract the tendency to produce false alarms. Nevertheless, CM could support other control functions related to errors. For instance, Botvinick et al. (2001) proposed that the adjustment of performance following an error, called *posterror slowing*, is driven by CM. Indeed, Rabbitt (2002) suggested that posterror slowing occurs even in the absence of conscious error detection. Thus, whereas RM-based error detection could be the mechanism underlying the ESR, CM-based error detection could support the adjustment of behavior following errors.

References

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. (1995). “Where did I go wrong?” A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1312–1322.

Bogacz, R., & Cohen, J. D. (2004). Parameterization of connectionist models. *Behavior Research Methods, Instruments, & Computers, 36,* 732–741.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

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.

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

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica, 101,* 379–394.

Endrass, T., Franke, C., & Kathmann, N. (2005). Error awareness in a saccade countermanding task. *Journal of Psychophysiology, 19,* 275–280.

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

Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1995). Event-related potential correlates of errors in reaction tasks. *Electroencephalography and Clinical Neurophysiology Supplement, 44,* 287–296.

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 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.

Fiehler, K., Ullsperger, M., & Von Cramon, D. Y. (2005). Electrophysiological correlates of error correction. *Psychophysiology, 42,* 72–82.

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

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

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General, 134,* 163–191.

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

Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica, 43,* 199–224.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

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

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38,* 752–760.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 358–377.

Rabbitt, P. (1966a, October 22). Error correction time without external error signals. *Nature, 212,* 438.

Rabbitt, P. (1966b). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology, 71,* 264–272.

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

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

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

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

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

Rabbitt, P., & Phillips, S. (1967). Error-detection and correction latencies as a function of S-R compatibility. *Quarterly Journal of Experimental Psychology, 19,* 37–42.

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

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9,* 347–356.

Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research, 66,* 312–323.

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.

Steinhauser, M., & Hübner, R. (2006). Response-based strengthening in task shifting: Evidence from shift effects produced by errors. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 517–534.

Ullsperger, M., & von Cramon, D. Y. (2006). How does error correction differ from error signaling? An event-related potential study. *Brain Research, 1105,* 102–109.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111,* 939–959.

Appendix A

The Basic Model

Here, we describe the details of the models used to simulate the response monitoring (RM) and conflict monitoring (CM) accounts of error detection. The formulas specifying the model dynamics in the task module and the CM device correspond to those of Yeung et al. (2004). The architecture of the task module was modified to match our paradigm, as described in the main text. We begin by describing the task module. Then, we report the extensions for the RM account and the CM account. The parameter values used by Yeung et al. (2004) are presented in Table A1.

The task module consists of a stimulus layer (30 units), a response layer (3 units), and an attention layer (3 units). The layers are connected as described in Figure 1. There is a constant associative weight w_{SR} for the feedforward excitatory connections between the stimulus and response layer and a weight w_{SA} for the bidirectional connections between the stimulus and attention layer. Furthermore, each unit inhibits each other unit within the same layer. This is achieved by negative weights w_S, w_A, and w_R for the stimulus, the attention, and the response layer, respectively.

In each processing cycle n, the network was updated in two steps. In a first step, the net input for each unit i was calculated as follows:

$$net_i(n) = [ext_i(n) \cdot estr] + \sum_j act_j(n-1)w_{ij}sc + noise$$

Table A1
Parameter Values From Yeung et al.'s (2004) Original Model

Parameter	Value
w_{SR}	1.5
w_{SA}	2.0
w_S	−2.0
w_A	−1.0
w_R	−3.0
$estr$	0.4
sc_e	0.08
sc_i	0.12
s_{noise}	0.035
α	4.41
β	1.08
γ	0.5
$decay$	0.1
act_{min}	−0.2
act_{max}	1.0
act_{rest}	−0.1
ext_S	0.15
ext_R	0.03
C	0.18
d_{stop}	6.0
s_{stop}	0.5
T_{cycle}	16
T_{ND1}	200

where $ext_i(n)$ is the external input at cycle n, $estr$ is a scaling parameter, w_{ij} is the weight of the incoming association from unit j, $act_j(n - 1)$ is the activation of this unit j on the preceding cycle, and sc is a further scaling parameter that is different for excitatory (sc_e) and inhibitory input (sc_i). Finally, there is normally distributed noise, taken from a distribution with mean 0 and standard deviation s_{noise}. Whereas the external input to the units in the stimulus and response layer was constant across cycles, that of the attention units was calculated by

$$ext_C(n) = \lambda ext_C(n - 1) + (1 - \lambda)[\alpha E(n - 1) + \beta]$$

and

$$ext_L(n) = ext_R(n) = [3 - ext_C(n)]/2,$$

where α, β, and γ are constants, and $E(n - 1)$ represents the response conflict on the preceding cycle. The value for ext_C was bound to the interval [1; 3].

In a second step, the activation change $\Delta act(n)$ for each unit i was determined by

$$\Delta act_i(n) = \{[act_{max} - act_i(n - 1)] \cdot net_i\} - \{[act_i(n - 1) - act_{rest}] \cdot decay\}$$

for $net_i(n) > 0$ and by

$$\Delta act_i(n) = \{[act_i(n - 1) - act_{min}] \cdot net_i\} - \{[act_i(n - 1) - act_{rest}] \cdot decay\}$$

for $net_i(n) < 0$.

In these formulas, $act_i(n - 1)$ is the activation in the preceding cycle, and the terms act_{min}, act_{max}, and act_{rest} represent the minimum, the maximum, and the resting activation, respectively. Finally, $net_i(n)$ is the unit's net input in this cycle, and $decay$ is a decay parameter. On the basis of the activations in the response layer, response conflict E in each cycle n was computed by

$$E(n) = -\sum_i \sum_j act_i(n) act_j(n) w_{ij}$$

where i and j denote each pair of units in the response layer. A product of unit activations was set to 0 whenever one of the activations was negative. This is part of the present model, because otherwise the conflict measure would have produced implausible values. In our simulations, the response conflict was computed not only for the CM model but also for the RM model to update the external input for the attention units. In this way, the task module behaved similarly for both models. However, the results reported in the present study do not depend on whether this mechanism was implemented.

Simulation Details

Each trial was simulated by a constant number of 50 cycles. During the first three cycles, external input was given only to the response units. In the fourth cycle, ext_i was also initialized for units in the stimulus and attention layer. In the stimulus layer, only the stimulus units representing the present flanker display received external input. A response was selected whenever the activation of the corresponding response unit exceeded the response criterion C. When this happened for the first time on a trial, the external input to all units was stopped after a random number of cycles, which was normally distributed with mean d_{stop} and standard deviation s_{stop}. If a second response unit exceeded the threshold, it was considered as an internal correction response.

To fit the model to the data, we had to transform the cycles into response times. For the first response, the corresponding latency RT_{first} was calculated by

$$RT_{first} = T_{cycle} \cdot n_{first} + T_{ND1}.$$

In this formula, n_{first} denotes the cycle at which the first response is selected. T_{cycle} represents the duration of a cycle in milliseconds, and T_{ND1} is a nondecisional time constant, which includes the duration of perceptual and motor processes related to the production of the first overt response. In the model of Yeung et al. (2004), T_{cycle} and T_{ND1} were set to 16 ms and 200 ms, respectively. In our study, these parameters were estimated from the data to obtain an optimal fit.

The latency of an overt correction response, error correction response (ECR), was calculated by

$$RT_{ECR} = T_{cycle} \cdot (n_{corr} - n_{first}) + T_{ND2}.$$

Here, n_{corr} denotes the cycle at which the internal response unit exceeds the threshold for the second time, and T_{ND2} is a nondecisional component comprising processes related to the initiation and execution of the overt correction response. It was assumed that the probability of ECRs equals the probability of internal correction responses—that is, $P(\text{ECR}) = P(\text{corr})$.

RM Model

To simulate the RM model, we simply computed the latency and frequency of an error signaling response for response monitoring (ESR_{RM}). The latency of this response depends on the duration of the internal correction response and on a nondecisional component T_{ND3}—that is,

$$RT_{ESR\text{-}RM} = T_{cycle} \cdot (n_{corr} - n_{first}) + T_{ND3}.$$

Note that the only difference between the formulas for the ESR_{RM} and that for the ECR is the nondecisional component. Here, the component T_{ND3} represents the duration of processes related to the initiation and execution of the ESR_{RM}. The probability of an ESR_{RM} was estimated by $P(\text{ESR}_{\text{RM}}) = P(\text{corr})[1 - P(\text{ESR fails})]$, where P(ESR fails) denotes the probability that the initiation and execution of an ESR fails despite an internal correction response having occurred.

CM Model

To simulate the CM model, we computed the cumulated conflict by

$$E_{cum}(n) = E_{cum}(n - 1) + E(n)$$

for $n \geq n_{first} + D$ and by

$$E_{cum}(n) = 0$$

for $n < n_{first} + D$,

(*Appendixes continue*)

where D is the delay after which conflict accumulation starts following the initial response. When the cumulated conflict exceeded the detection threshold K, an error was detected. The latency of the corresponding $\mathrm{ESR_{CM}}$ is given by

$$RT_{\text{ESR-CM}} = T_{\text{cycle}} \cdot (n_{\text{det}} - n_{\text{first}}) + T_{\text{ND3}},$$

where n_{det} denotes the time at which the cumulated conflict exceeded the threshold. T_{ND3} represents the nondecisional component comprising the initiation and execution of the $\mathrm{ESR_{CM}}$ (which is similar to that of the $\mathrm{ESR_{RM}}$). The estimated probability of an $\mathrm{ESR_{CM}}$ equals the frequency that an error is detected—that is, $P(\mathrm{ESR_{CM}}) = P(\text{det})$.

Appendix B

Simulation Experiment

To see how the response criterion affects the error correction response (ECR) and the error signaling response (ESR) performance in the response monitoring (RM) and conflict monitoring (CM) models, we conducted a simulation experiment. If not mentioned otherwise, we used the same parameters as Yeung et al. (2004) in their two-choice model. Note that these parameters were not estimated by fitting the model to data. Rather, they were chosen by Yeung et al. and Botvinick et al. (2001) because they produced qualitatively plausible results. For our models, these parameters also produced plausible results, although we used a slightly different architecture.

For each model, we simulated 5,000 trials for neutral and incongruent stimuli and five response criteria (from 0.14 to 0.22 in steps of 0.02). Because we were not interested in the effect of congruency, the data were collapsed for this variable. This is unproblematic as the obtained results were rather similar for neutral and incongruent stimuli. Thus, each conclusion in the following holds for both stimulus types. We also conducted the same simulations with the original two-response model, using identical parameters. Both models produced nearly the same pattern. Therefore, we restrict our consideration to the three-response models.

Because criterion effects can emerge only for the decisional part of our models, we calculated only the performance of this component, measuring time by means of model cycles n. More specifically, the latency of correct initial responses was calculated by the cycle a first response unit exceeded the threshold, n_{first}, for correct responses. The latencies of the ECR (in both models) and the $\mathrm{ESR_{RM}}$ (in the RM model) were all calculated by the internal correction time, $n_{\text{corr}} - n_{\text{first}}$. Finally, the latency of the $\mathrm{ESR_{CM}}$ (in the CM model) was calculated by the time until the CM-based detector detects an error, $n_{\text{det}} - n_{\text{first}}$. Similarly, the overall error rate was calculated as the frequency of first erroneous response, the probability of an ECR and an $\mathrm{ESR_{RM}}$ was calculated as the frequency of internal correction responses, and the probability of an $\mathrm{ESR_{CM}}$ was calculated as the frequency of error detections by the CM-based error detector. From this, it already becomes obvious that response criterion effects are by definition similar for the ECR and the $\mathrm{ESR_{RM}}$. Please note that, using this method, we can illustrate the effect of response criterion on the various measures, but we cannot compare their absolute performance, which additionally relies on nondecisional components.

As already mentioned, the CM model requires two additional parameters: one parameter for the delay after which conflict accumulation starts and a detection threshold. Because Yeung et al. (2004) have shown that a delay of six cycles is appropriate, we used the same value. However, as becomes clear below, for our objective it was also necessary to simulate the performance with a delay of 0. Moreover, we used two different detection thresholds of 0.05 and 0.0001 to demonstrate the outcome of a conservative and a liberal detector, respectively.

Results and Discussion

The results of our simulations are shown in Figure B1, where each set of connected points represents the criterion effect on one dependent variable. In the following, we consider each result separately.

Correct Initial Responses

Inspection of the first column shows that response times of correct responses (see Figure B1A) increased with an increasing criterion, whereas the error rates decreased (see Figure B1B). This pattern reflects the speed–accuracy trade-off, which is usually observed when the response criterion is manipulated. It confirms that the model behaved as expected.

ECR/ESR$_{RM}$

The second column shows the simulated ECR and $\mathrm{ESR_{RM}}$ performance. As expected, the model exhibited a strong criterion effect on the latency for both measures (see Figure B1C). This is attributable to the fact that an internal correction response occurs only when the activation of the correct response exceeds the criterion, which requires more time when the criterion is higher (as illustrated in Figure 3 in the main text). Interestingly, the effect of the criterion on ECR and $\mathrm{ESR_{RM}}$ latency was even stronger than the corresponding effect on the response time for correct responses. This is mainly because of the distribution of response activation at the cycle when the initial response is selected. At this time, a higher criterion implies that the wrong response is activated more strongly, reflecting the higher threshold that had to be exceeded. As a consequence, a higher criterion requires that the internal correction response needs even more time to overcome the erroneous response.

As can be seen in Figure B1D, the estimated probability of an ECR and $\mathrm{ESR_{RM}}$ decreased with an increasing criterion. Basically, this is the result of the interrupted stimulus processing after the first response. The correction response not only requires more activation to reach the increased criterion, it is also no longer activated by the stimulus. Inspection of Figure B1D also makes it

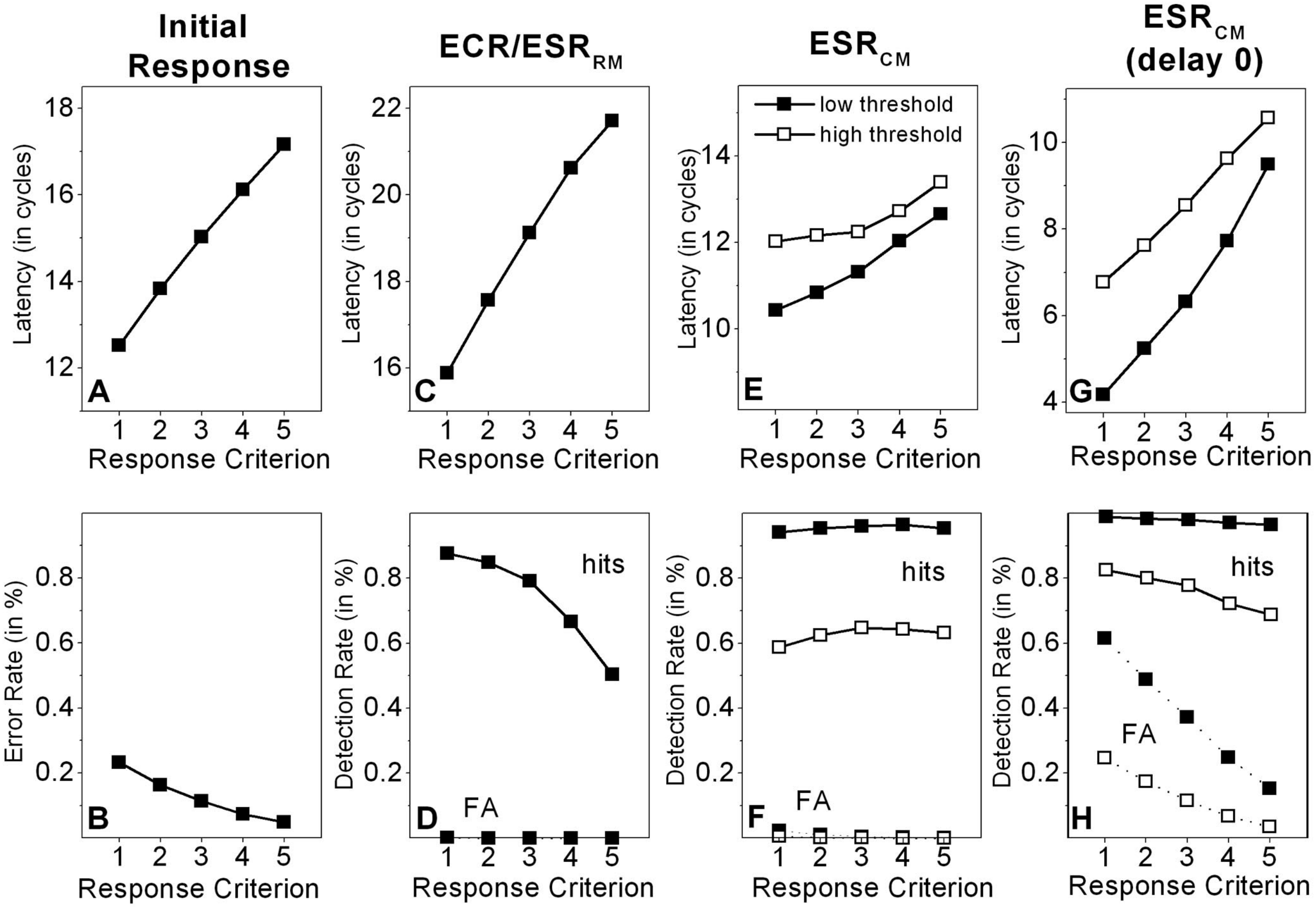

Figure B1. Simulation results. Upper row: Latencies of the initial response on correct trials (A), the ECR/ESR$_{RM}$ (C), the ESR$_{CM}$ (E), and the ESR$_{CM}$ with a delay of 0 (G). Lower row: Frequencies of errors (B) and frequencies of hits and false alarms (FAs) on ECR/ESR$_{RM}$ (D), ESR$_{CM}$ (F), and ESR$_{CM}$ with a delay of 0 (H). For the ESR$_{CM}$, all data are shown for a low and a high detection threshold. ECR = error corrections response; ESR$_{RM}$ = error signaling response by response monitoring; ESR$_{CM}$ = error signaling response by conflict monitoring.

obvious that false alarms, which happen when a correct response is internally corrected by an erroneous response, were rare. If at all, they occurred for very low criteria, where small fluctuations in response activation due to noise can exceed the criterion.

Altogether, our simulations demonstrated that an error detector based on response monitoring is very reliable. It signals a negligible rate of false alarms, and the number of misses is high only when stimulus processing ends early. With a sufficiently long period of stimulus processing, all errors should be corrected and, therefore, detected.

ESR$_{CM}$

The third column of Figure B1 shows the results for the CM-based error detector. Graphs are provided for both a high and a low detection threshold. In contrast to our expectation, there is a criterion effect on ESR$_{CM}$ latency (see Figure A1E in Appendix A). However, relative to the criterion effect on ECR latency, the slopes are small. Apparently, the response criterion also has a much smaller effect on ESR$_{CM}$ latency than it does on the latency of the initial response. Furthermore, the hit rate function has a slightly positive slope for the smaller criteria and a slightly negative slope for the higher response criteria. The false alarm rate is, as expected, rather low.

Our simulation revealed that even the CM-based error detector predicts a small but systematic effect of the response criterion on detection performance. How this effect could emerge is illustrated in Figure 4 in the main text. At the time the error has occurred, the activation difference between the correct and the erroneous responses is larger with a higher criterion. As a consequence, the response conflict at this time is lower with a higher criterion. After some cycles, the criterion effect on response conflict largely disappears. However, the initial influence is sufficient for having a slight effect on cumulated response conflict, on which error detection is based, and which starts after six cycles. For the low criterion, the buildup of cumulated conflict is slightly steeper in the first cycles, which leads to an earlier exceeding of the detection threshold. However, the asymptote toward which the cumulated conflict converges is slightly higher with a high criterion, which explains why the frequency of detected errors increases with the criterion.

(*Appendixes continue*)

Apparently, a criterion effect on the $\mathrm{ESR}_{\mathrm{CM}}$ is obtained when conflict accumulation starts early enough for being affected by the early conflict difference between low and high response criteria. To test this, we also computed the results for a CM-based error detector when conflict accumulation starts with a 0 delay. When conflict accumulation starts immediately after the erroneous response, there is a large effect of response criterion on detection performance, as revealed in the fourth column of Figure B1. For this case, the performance of the CM-based detector is rather similar to the performance of the RM-based error detector. However, there is one difference. With a 0 delay, the CM-based error detector produces a high number of false alarms, especially when the detection threshold is low (see Figure B1H). Since this is typically not observed, small delays should be inappropriate for modeling empirical data. As already mentioned, Yeung et al. (2004) delayed the onset of measuring the cumulated conflict to prevent the production of too many false alarms.

Conclusions

The goal of these initial simulations was to clarify how each account predicts the influence of the response criterion on detection performance. As expected, we found a strong influence of response criterion on ECR or $\mathrm{ESR}_{\mathrm{RM}}$ performance. However, we also found an effect of response criterion on $\mathrm{ESR}_{\mathrm{CM}}$ performance. This latter effect is mainly attributable to the fact that conditions with high and low response criteria differ with respect to response conflict in the first posterror cycles. Consequently, the shorter the delay after which conflict measurement is started, the stronger the effect of response criterion on $\mathrm{ESR}_{\mathrm{CM}}$ performance. Although short delays are implausible because they also produce high false alarm rates, a small criterion effect is obtained even with a long delay.

The simulation suggests that the two models differ mainly with respect to the size of the criterion effect on ESR latency. Whereas the RM account implies that the criterion effect on ESR latency should be identical to that on ECR latency, the CM account would predict a much smaller effect on ESR latency than on ECR latency.

Received April 21, 2006
Revision received October 25, 2006
Accepted December 29, 2006 ■