Exp Brain Res (2010) 202:571–581
DOI 10.1007/s00221-010-2160-3

RESEARCH ARTICLE

Effects of response-set size on error-related brain activity

Martin E. Maier · Marco Steinhauser · Ronald Hübner

Received: 12 August 2009 / Accepted: 7 January 2010 / Published online: 22 January 2010

Abstract To study the effect of response-set size on action monitoring processes, the error-related negativity (Ne/ERN), the correct-related negativity (Nc/CRN), and behavioral indicators of action monitoring were compared across three groups of participants performing a two-choice, a four-choice, or an eight-choice version of the flanker task. After controlling for differential contribution of stimulus-related activity to response-locked averages resulting from large differences in response times across conditions, response-set size had strong effects on Ne/ERN and Nc/CRN. With increasing response-set size, the Ne/ERN amplitude decreased, but the Nc/CRN amplitude increased. Moreover, post-error behavioral adjustments were impaired with an increasing response-set size. These results suggest that action monitoring severely suffers when response-set size is increased. Implications of these findings for present theories of Ne/ERN and Nc/CRN are discussed.

Keywords Action monitoring · Error-related negativity · Response uncertainty · Post-error slowing · Error detection · Error signaling response

Introduction

Action monitoring is crucial for goal-directed behavior and the optimization of performance. In recent years, the so-called error negativity (Ne) or error-related negativity (ERN) has been established as an important measure of action monitoring. The Ne/ERN is a negative deflection in the event-related potential (ERP) peaking shortly after erroneous responses on fronto-central channels (Falkenstein et al. 1990; Gehring et al. 1993) and is presumed to be generated in the anterior cingulate cortex (ACC, Carter et al. 1998; Dehaene et al. 1994; Ullsperger and von Cramon 2001). In addition to the Ne/ERN on error trials, a smaller negativity with similar latency and scalp distribution was observed on correct trials, the correct negativity (Nc) or correct-related negativity (CRN) (e.g., Falkenstein et al. 2000; Luu et al. 2000; Vidal et al. 2000).

Several theories have been proposed assuming a direct relation between the Ne/ERN and action monitoring. Whereas the mismatch hypothesis posits that the Ne/ERN represents the amount of mismatch between the actual response and the correct response (e.g., Bernstein et al. 1995; Falkenstein et al. 2000), conflict monitoring theory suggests that it reflects a conflict between the error response and the later activated correct response (Yeung et al. 2004). Both accounts imply that the detection of a mismatch/conflict enables error detection. Within these frameworks, the Nc/CRN could be assumed to reflect erroneous mismatch or conflict on correct trials that lead to false alarms (e.g., Coles et al. 2001; Vidal et al. 2000).

According to these accounts, the generation of an Ne/ERN crucially depends on whether the action monitoring system has an intact representation of the correct response. Whatever impairs this correct response representation should impair the Ne/ERN. Therefore, a simple strategy to test these accounts is to investigate the relation between the Ne/ERN and variables affecting the efficiency by which the correct response is derived. To identify such variables, one has to consider how the action monitoring system determines the correct response. For instance, conflict monitoring theory (Yeung et al. 2004) assumes that the response

M. E. Maier (✉) · M. Steinhauser · R. Hübner
Fachbereich Psychologie, Universität Konstanz,
Fach D29, 78457 Constance, Germany
e-mail: martin.maier@uni-konstanz.de

selection mechanism producing the error has an inherent tendency to immediately correct the error. After an error has occurred, the correct response becomes activated due to continued stimulus processing and elicits a conflict with the still active error response, and this conflict is reflected in the Ne/ERN. Thus, the same mechanism that normally derives a correct response is responsible for deriving the correction response in case of an error. This idea leads to the general prediction that any variable impairing response selection also impairs the corrective tendency that gives rise to post-error conflict and the Ne/ERN. Moreover, such a variable should also impair error detection, because error detection is assumed to rely on the detection of post-error conflict (Yeung et al., 2004; but see Steinhauser et al. 2008). A similar reasoning is applicable if we assume a mismatch between the error response and the correct response instead of post-error conflict to be the source of the Ne/ERN.

In support of this prediction, a number of variables have been identified that simultaneously impair task performance as well as the Ne/ERN, such as target-distractor congruency (Danielmeier et al. 2009; Yeung et al. 2004), or time pressure (Gehring et al. 1993). Recently, however, an interesting exception to this rule has been reported. Pailing and Segalowitz (2004) examined the Ne/ERN as a function of different variables known to impair task performance. In one experiment, stimulus discriminability in a tone categorization task was manipulated. With reduced discriminability, errors were associated with a smaller Ne/ERN, whereas correct responses were associated with a larger Nc/CRN. This can be explained by assuming that less discriminable stimuli not only impair the activation of the correct response, but also the activation of the corrective tendency after an error, and that this leads to lower post-response conflict and thus to a smaller Ne/ERN. Moreover, less discriminable stimuli increase the risk that the wrong response is activated following a correct response, which can explain why a larger Nc/CRN is observed. In a further experiment, Pailing and Segalowitz (2004) manipulated response-set size which produced no effect on Ne/ERNs or Nc/CRNs. The authors concluded that only stimulus discriminability but not response-set size impairs the representation of the correct response.

The absence of an effect of response-set size on the Ne/ERN seems to violate the general prediction that variables affecting response selection should also affect action monitoring. It is well known that response times and error rates in choice tasks increase logarithmically with an increasing response-set size (Hick 1952). These effects have frequently been attributed to an increase in response uncertainty. Increasing the number of possible responses increases the probability that the stimulus erroneously activates an incorrect response which slows responding and/or causes errors (e.g., Usher et al. 2002). As a consequence, response uncertainty should also impair the corrective tendency after an error. Accordingly, error detection should be more difficult under these conditions. And indeed, there is evidence that the ability to consciously report errors is impaired with an increasing response-set size (Rabbitt 1967).

Thus, the question emerges why Pailing and Segalowitz (2004) did not find an effect of response-set size on the Ne/ERN. The answer could be that their manipulation was simply not strong enough. Pailing and Segalowitz (2004) compared a condition with two response alternatives and a condition with three response alternatives. Indeed, studies investigating Hick's Law typically compare response sets of size two, four, and eight (e.g., Alegria and Bertelson 1970). In the present work, the question whether response-set size affects action monitoring was put to a further test. To obtain a strong manipulation of response-set size, the data of participants performing a two-choice version, a four-choice version, or an eight-choice version of the flanker task (Eriksen and Eriksen 1974) were compared. Following Rabbitt (1967), a between-subjects design was chosen so that participants did not have to change response sets and mappings during the experiment in order to avoid interference. Based on the findings of Pailing and Segalowitz (2004), it was hypothesized that the Ne/ERN amplitude on error trials decreases, but the Nc/CRN on correct trials increases with an increasing response-set size.

In addition to error-related brain activity, a number of behavioral indicators of action monitoring were considered. First, error detectability was measured by requiring the participants to signal their errors immediately (e.g., Rabbitt 2002). Second, post-error slowing was analyzed which refers to a slowing of response times following errors and which is viewed as a behavioral adjustment that prevents further errors (Laming 1979; Rabbitt 1966). It was hypothesized that error detectability and post-error slowing would also be impaired with a larger response-set size. Such a pattern of results would suggest that action monitoring processes are more susceptible to failure with larger response-set sizes.

Methods

Participants

Forty-two participants were randomly assigned to three groups of 14 participants each. A first group (10 female, mean age 22.3) performed the two-choice task, a second group (10 female, mean age 23.2) performed the four-choice task, and a third group (11 female, mean age 22.9) performed the eight-choice task. All participants had normal or corrected-to-normal vision, were recruited at the

University of Konstanz, and received 5 Euro/h. The study was conducted in accordance with institutional guidelines and informed consent was acquired from all participants.

Stimuli and tasks

Stimuli were presented on a 21-in. color monitor in white on a black background. They were composed of the letters B, K, P, R, M, V, W, or X, and of the neutral symbols %, &, ?, #, $, §, or ¥ taken from the Arial font. Each character was resized to a visual angle of 0.64° height and 0.48° width at a viewing distance of 72 cm. Each stimulus array consisted of a central target letter flanked by three identical distractors on each side which were either letters or symbols. The whole array subtended a visual angle of 4.2° width.

The three tasks were constructed by mapping the same eight letters on varying numbers of responses. A 4:1 mapping was used in the *two-choice task* (letters B, K, W, X were assigned to one response, and letters M, V, P, R were assigned to another response), a 2:1 mapping was used in the *four-choice task* (the pairs B and K, P and R, M and V, W and X were assigned to one of four responses each)[1], and a 1:1 mapping was used in the *eight-choice task* (each letter was assigned to one of eight responses). Within each task, incongruent stimuli were constructed by combining each possible target with the letters associated with a different response as the target. This resulted in 32 incongruent stimuli in the two-choice task, 48 incongruent stimuli in the four-choice task, and 56 incongruent stimuli in the eight-choice task. Neutral stimuli were constructed by combining each target in the two-choice task with one of the symbols %, &, §, or # (32 neutral stimuli), each target in the four-choice task with one of the symbols $, ?, %, &, §, or # (48 neutral stimuli), and each target in the eight-choice task with one of the symbols ¥, $, ?, %, &, §, or # (56 neutral stimuli).

Participants were instructed to respond to the target but to ignore the flankers. Responses were given using a German standard keyboard. Participants in the two-choice task had to press the ‘S’ and ‘L’ keys with the left and the right index finger, respectively. Participants in the four-choice task had to press the ‘W’, ‘S’, ‘L’, and ‘P’ keys with the left middle finger, the left index finger, the right index finger, and the right middle finger, respectively. Participants in the eight-choice task, had to press the ‘Q’, ‘W’, ‘E’, ‘D’, and ‘L’, ‘P’, ‘Ü’, ‘+’ keys with the left pinky, the left middle finger, the left index finger, the right pinky, the right index finger, and the right middle finger, respectively.

Procedure

Each trial started with the presentation of a fixation cross for 250 ms. Then, the stimulus array was presented for 150 ms followed by a blank screen. A new trial started after 1,200 ms. If further responses occurred during this interval, the interval was restarted. In some blocks, participants were instructed to give an error signaling response immediately whenever they detected an error by simultaneously pressing the ‘Alt’ key on the left side and ‘Alt-Gr’ key on the right side of the keyboard with the left and with the right thumb, respectively. The order of stimuli was randomized. Stimulus presentation and response registration was controlled using custom C++ routines.

To achieve similar levels of expertise for the different tasks, each stimulus was presented at an approximately similar frequency for each task. This was done because research on automaticity suggests that expertise does not depend on the time spent on each task but rather on the number of encounters with each stimulus (Logan 1988). However, this implies that the total number of trials differs for the three tasks because different numbers of stimuli were used for each task. In the two-choice task, 20 blocks of 64 trials each (total 1,280 trials), in the four-choice task, 16 test blocks of 96 trials each (total 1,536), and in the eight-choice task, 16 test blocks of 112 trials each (total 1,792) were administered. Blocks with signaling instruction and blocks without signaling instruction alternated. Half of the participants began with a block with signaling instruction and the other half with a block without signaling instruction.

The blocks were distributed over two test sessions lasting approximately 45 min in the two-choice task, 1 h in the four-choice task, and 1 h 15 min in the eight-choice task. In a preliminary practice session, participants had to perform eight practice blocks during which they learned the stimulus–response mapping. Furthermore, these blocks were used to adjust the error rate. Whenever the average error rate in a block fell below 15%, participants were instructed to respond faster at the beginning of the next block. This instruction was maintained throughout the whole experiment to achieve similar error rates across conditions. After six practice blocks without signaling instruction, subjects had to perform two practice blocks with signaling instruction.

Psychophysiological recording

Participants were seated comfortably in a dimly lit, electrically shielded room. The electroencephalogram (EEG) was

[1] The data from the four-choice task were already published elsewhere (Maier et al. 2008). However, the data were re-analyzed in a way that was appropriate for all three response-set size conditions. Accordingly, trimming of the behavioral data, filtering, baseline correction, and quantification of ERP components have changed.

recorded with Ag/AgCl electrodes mounted in a cap (Easycap, Herrsching, Germany) from three electrode sites: frontal (Fz), fronto-central (FCz), and central (Cz). The right mastoid was recorded as an additional channel. Electrodes were referenced to the left mastoid and off-line re-referenced to the average of both mastoids. Electrode impedances were kept below 5 kΩ. Vertical and horizontal electrooculogram (EOG) was recorded from above and below the left eye and from the outer canthi of both eyes, respectively. EEG and EOG were continuously recorded at a sampling rate of 200 Hz and a high-pass filter of 0.1 Hz using Biopac amplifiers (BIOPAC Systems, Goleta, CA, USA). Waveforms were off-line filtered with a low-pass filter of 40 Hz and a high-pass filter of 1 Hz.

Data analysis

Data analyses were done using MatLab 7.0.4 (The Mathworks, Natic, MA, USA) and EEGLAB 5.03 (Delorme and Makeig 2004), an open source toolbox for EEG data analysis (EEGLAB toolbox for single-trial EEG data analysis, Swartz Center for Computational Neurosciences, La Jolla, CA; http://www.sccn.ucsd.edu/eeglab).

Behavioral data

Response time was defined as the time interval between the onset of the stimulus and the subsequent key press. Error trials on which both signaling keys were pressed were classified as trials with valid signaling responses. The latency of a valid signaling response was calculated as the mean interval between the erroneous response and each of the two signaling keys. Post-error slowing was computed as the response time of correct responses on trials following an error minus the response time of correct responses on trials following another correct response.

To control for outliers, trials were excluded for which the response time of the choice response was three standard deviations above or below the condition mean (2.7% in the two-choice task, 2.5% in the four-choice task, and 2.0% in the eight-choice task). Furthermore, trials were excluded on which only one signaling button was pressed (<1% in all three tasks). Finally, trials with a spontaneous error correction were excluded (9.5% of all errors in two-choice task, 8.8% of all errors in the four-choice task, 7.3% of all errors in the eight-choice task).

Signaling latencies and frequencies were subjected to two-way ANOVAs with the variables' response-set size (two, four, eight), and congruency (neutral, incongruent). All other response time and error rate data were subjected to three-way ANOVAs on the variable response-set size (two, four, eight), block type (with signaling instruction, without signaling instruction), and congruency (neutral, incongruent). The variable block type was included in the experimental design, because the data from the four-choice task were used in a study published elsewhere, which required blocks without signaling instruction (see, Maier et al. 2008). Here, this is not of importance and therefore, results concerning the variable block type are not reported.

ERP data

Epochs of 500 ms before and 1,000 ms after the first response were extracted from the continuous EEG. The average voltage in a window ranging from 75 to 25 ms preceding the response served as a baseline, and was subtracted from each epoch[2]. Epochs contaminated with large artifacts were identified using three methods from the EEGLAB toolbox (see, Delorme et al. 2007). An epoch was excluded (1) whenever the voltage on an EOG channel exceeded an individually adjusted threshold to remove epochs with large EOG peaks, (2) whenever the joint probability of a trial exceeded five standard deviations to remove epochs with improbable data, and (3) whenever the kurtosis exceeded five standard deviations to remove epochs with unusually distributed data. The mean percentage of trials excluded was 31% and did not differ between the three tasks, $F < 1$. Remaining horizontal and vertical EOG artifacts were corrected by an eye movement correction procedure (Automatic Artifact Removal Toolbox Version 1.3; http://www.cs.tut.fi/~gomezher/projects/eeg/aar.htm) based on a linear regression method described by (Gratton et al. 1983). One participant in the two-choice task had to be excluded, because of excessive eye artifacts.

In this way, an average number of 169, 210, and 243 correct trials, and 46, 63, and 64 error trials per participant in the two-choice, the four-choice, and the eight-choice tasks, respectively, remained in the analyses. These ERP signals were filtered with infinite impulse response filters with low cutoff of 1 Hz and high cutoff of 10 Hz. A low-pass filter of 10 Hz was chosen to prevent high-frequency noise from distorting peak component measures. The data were then averaged locked to the response.

Ne/ERN and Nc/CRN were quantified using baseline-independent base-to-peak measures[3]. The Ne/ERN was

[2] This baseline window was chosen, because it aligned the waveforms for correct and error responses with respect to the positive peaks preceding error-related brain activity. This best illustrates differences in Nc/CRN and Ne/ERN across conditions. Note, however, that the choice of baseline did not affect statistical analyses, because baseline-independent base-to-peak measures were used for component quantification.

[3] We also analyzed the data using the mean voltage in a time window of −25 to 100 ms relative to the response for the Ne/ERN and −25 to 50 ms relative to the response for the Nc/CRN. This did not change the results qualitatively.

quantified as a the difference between the most positive peak in a time window of −100 to −25 ms relative to the response and the most negative peak in a time window of −25 to 100 ms relative to the response. The Nc/CRN was quantified as a the difference between the most positive peak in a time window of −100 to −25 ms relative to the response and the most negative peak in a time window of −25 to 50 ms relative to the response. Analyses were conducted using data from electrode Fz because both the Ne/ERN and the Nc/CRN showed maximum amplitudes at this electrode site. Because we predicted different effects of response-set size on Ne/ERN and Nc/CRN, we intended to analyze the data using a three-way ANOVA with the variables component (Ne/ERN, Nc/CRN), response-set size (two, four, eight), and congruency (neutral, incongruent). Where necessary, post hoc testing of directional hypotheses was done using one-tailed *t* tests. Using one-tailed *t* tests was justified because we tested between the pattern predicted by Pailing and Segalowitz (2004) and no effect (or random effects in other directions), which constitutes a strong directional hypothesis.

To examine the contribution of stimulus-related activity on the response-locked averages, ERP distributions were investigated using a running averages procedure described below. Because this analysis revealed that our main conditions differed with respect to the contribution of the stimulus-locked P300, the data were re-analyzed after filtering with a 4 Hz infinite impulse response high-pass filter to remove this component (Vidal et al. 2003).

For completeness, we also computed difference waves between errors and correct trials. For each participant, the maximum difference in the interval −25 to 100 ms relative to the response was determined, and these data were entered into a two-way ANOVA with the variables response-set size (two, four, eight), and congruency (neutral, incongruent).

Results

Behavioral data

The behavioral data are presented in Table 1. Whereas error rates were similar across response-set sizes (two, 20.7%; four, 20.1%; eight, 19.4%), a significant interaction between response-set size and congruency, $F(2, 39) = 6.18$, $p < 0.01$, was obtained indicating that the congruency effect (incongruent − neutral) was decreased with an increasing response-set size (two, 6.4%; four, 4.4%; eight, 2.6%). In contrast to the error rates, strong effects of response-set size were present in response times of correct responses (two, 480 ms; four, 558 ms; eight, 634 ms), $F(2, 39) = 51.8$, $p < 0.001$, as well as in response times of error responses (two, 424 ms; four, 528 ms; eight, 635 ms), $F(2, 39) = 47.9$, $p < 0.001$. Furthermore, reliable congruency effects were obtained for correct response times, $F(1, 39) = 159$, $p < 0.001$, and error response times, $F(1, 39) = 12.4$, $p < 0.001$.

The signaling latencies showed a strong trend towards a main effect of response-set size (two, 438 ms; four, 457 ms; eight, 518 ms), $F(2, 39) = 2.70$, $p = 0.08$, but no such effect was found for the frequencies of signaling responses (two, 97.1%; four, 95.4%; eight, 95.2%), or the frequencies of false alarms (two, 0.82%; four, 0.79%; eight, 0.85%). Finally, there was a reliable effect of response-set size on post-error slowing indicating stronger post-error slowing in the two-choice task (19 ms) than in the four-choice task (8 ms) and the eight-choice task (6 ms), $F(1, 39) = 4.14$, $p < 0.05$. Post hoc tests showed that post-error slowing was stronger in the two-choice task (19 ms) than in the four-choice task (8 ms), $t(25) = 2.21$, $p < 0.05$, and in the eight-choice task (6 ms), $t(25) = 2.70$, $p < 0.01$. However, it did not differ significantly between the four-choice task and the eight-choice task. Taken together, the analyses of behavioral data show that an increasing response-set size impairs not only the selection of the initial response, as predicted by

Table 1 Behavioral data in the three versions of the flanker task

	Two-choice		Four-choice		Eight-choice	
	Neutral	Incongruent	Neutral	Incongruent	Neutral	Incongruent
% Errors	17.5 ± 1.0	23.9 ± 1.2	17.8 ± 0.9	22.5 ± 1.1	18.1 ± 0.6	20.7 ± 0.8
RT correct	474 ± 6.1	486 ± 6.2	550 ± 4.7	566 ± 5.4	625 ± 10.3	643 ± 10.4
RT error	420 ± 7.5	427 ± 7.0	526 ± 8.9	530 ± 10.3	627 ± 13.7	644 ± 15.7
% Errors signaled	97.0 ± 2.1	97.2 ± 2.0	95.0 ± 3.6	95.9 ± 2.2	94.8 ± 1.5	95.5 ± 1.3
Signaling latency	435 ± 24.4	440 ± 24.6	455 ± 26.9	459 ± 27.0	523 ± 24.9	513 ± 26.8
% False alarms	0.76 ± 0.18	0.89 ± 0.27	0.85 ± 0.25	0.72 ± 0.13	0.75 ± 0.17	0.94 ± 0.18
Post-error slowing	18 ± 4.2	21 ± 3.5	11 ± 2.7	6 ± 3.2	7 ± 4.7	5 ± 3.9

All latency measures are in milliseconds. Values are expressed as mean ± SE

RT response time

Hick's Law, it also impairs action monitoring as indicated by reduced post-error slowing and a trend towards longer signaling latencies.

ERP data

Base-to-peak analysis of average ERPs filtered with 1–10 Hz

The ERP data are presented in Fig. 1. Inspection of the waveforms in Fig. 1b revealed a clear effect of response-set size on the Ne/ERN. However, the Ne/ERN amplitude did not decrease monotonically with increasing response-set size. Rather, a larger Ne/ERN was obtained in the two-choice task (7.19 µV) as compared to the four-choice task (4.28 µV) or the eight-choice task (4.81 µV). Furthermore, inspection of the waveforms in Fig. 1a revealed a clear Nc/CRN only for the eight-choice task. Indeed, most participants did not show CRN peaks with smaller response-set sizes, what made a quantification of this component difficult. Because of this, only the Ne/ERN data were subjected to statistical analysis. A two-way ANOVA on the variables response-set size (two, four, eight) and congruency (neutral, incongruent) confirmed our initial impression by showing a significant main effect of response-set size, $F(2, 38) = 4.01$, $p < 0.05$. Moreover, post hoc tests showed that the Ne/ERN was larger in the two-choice task than in the four-choice task, $t(25) = 2.67$, $p < 0.05$, or the eight-choice task, $t(25) = 2.18$, $p < 0.05$, while it did not differ between the latter two conditions.

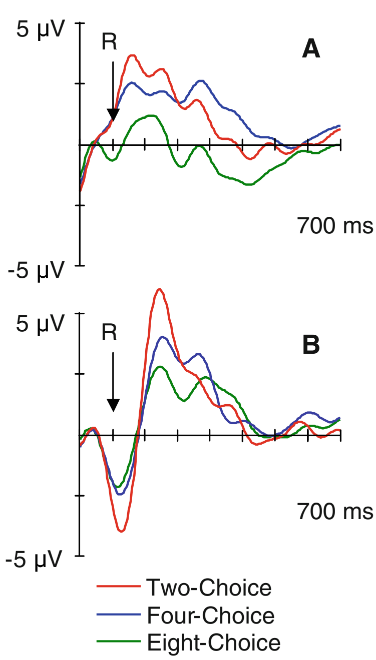

Fig. 1 Response-locked grand average waveforms filtered with a high-pass filter of 1 Hz and a low-pass filter of 10 Hz for correct trials (panel **a**) and error trials (panel **b**) in the two-choice task (*red lines*), in the four-choice task (*blue lines*), and in the eight-choice task (*green lines*), *ms* milliseconds, *µV* microvolt. See text for further details

A problem for our analysis could be the fact that response times vary considerably across response-set sizes. Because of this, response-related components and stimulus-related components overlap differentially in the post-response period which could lead to distortions (see also Coles et al. 2001; Hajcak et al. 2004). For instance, stimulus presentation typically causes a P300, i.e., a positivity emerging 300-600 ms after stimulus onset. Accordingly, due to the response times associated with the conditions, the post-response period for correct responses as well as for errors strongly overlaps with the P300 in the two-choice task (correct trials, 480 ms; error trials, 424 ms) and the four-choice task (correct trials, 558 ms; error trials, 528 ms) but not in the eight-choice task (correct trials, 634 ms; error trials, 635 ms). As a consequence, the post-response period should appear more positive in the two-choice task and the four-choice task due to superposition with stimulus-related activity. This implies an underestimation of the Nc/CRN and the Ne/ERN which could explain the absence of a clear Nc/CRN peak in these conditions as well as the non-monotonous pattern in the Ne/ERN. Thus, to obtain an unbiased measure of these components, the contribution of stimulus-locked components to our effects need to be isolated.

Distributional analysis

To explore whether our results are influenced by differential overlap of stimulus-locked and response-locked potentials in the three response-set sizes, we examined ERPs as a function of response time. To achieve this, a running averages procedure was applied as suggested by Coles et al. (2001). First, trials were sorted by response time for each condition and participant separately. Next, overlapping trial bins of 20 ms width were created, which were staggered by 5 ms (e.g., 300–320, 305–325, 310–330, etc.), and each trial was assigned to the respective bins. Then, average stimulus-locked ERPs were calculated for each bin containing more than two trials. Finally, grand averages of stimulus-locked ERPs were computed across participants for bins in which data were available for all participants and conditions. Figure 2 plots the results for correct and error trials collapsed across block type and congruency within each task at electrode site Fz. Note that in this figure, stimulus-related activity appears as vertical bands of voltage changes, and response-related activity appears as diagonal bands of voltage changes along the line representing the average response time in each bin.

Inspection of Fig. 2 reveals a strong positivity in the post-response period of correct trials with response times between 400 and 600 ms. This positive maximum presumably

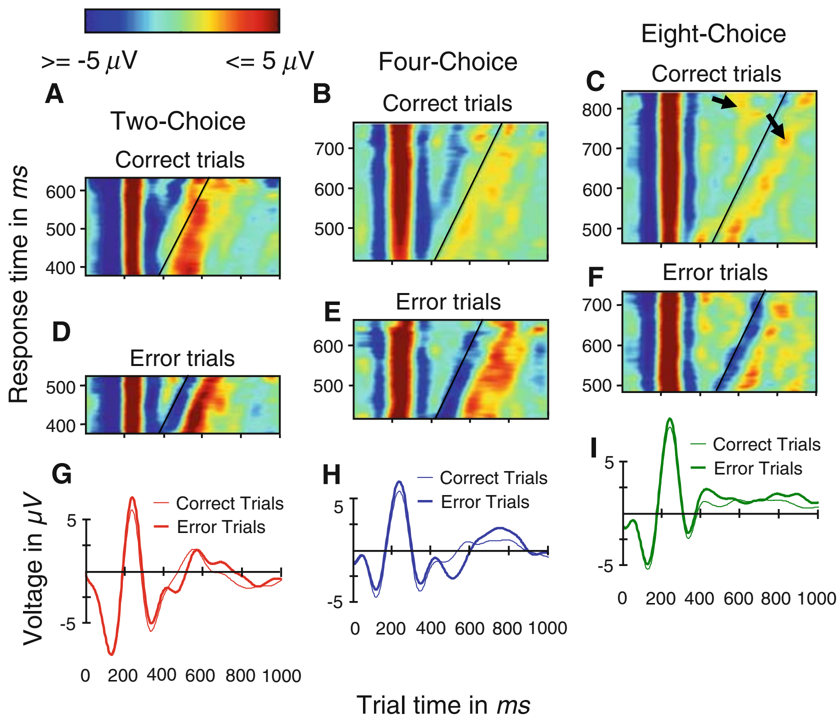

Fig. 2 Stimulus-locked voltage changes as a function of response time. ERPs were filtered by a 1 Hz high-pass filter and a 10 Hz low-pass filter, sorted by response time and assigned to overlapping response time bins of 20 ms width staggered by 5 ms. Averages for bins containing more than two trials were computed and plotted in a raster such that response time in milliseconds is represented on the ordinate, and trial time in milliseconds is represented on the abscissa. Positive voltages are depicted in *red*, and negative voltages are depicted in *blue*. Panels **a**, **b**, and **c** Correct trials in the two-choice task, the four-choice task, and the eight-choice task, respectively. Panels **d**, **e**, and **f** Error trials in the two-choice task, the four-choice task, and the eight-choice task, respectively. Panels **g**, **h**, and **i** Average stimulus-locked waveforms for correct trials (*thin lines*) and error trials (*thick lines*), respectively. The *diagonal black line* from the bottom to the top of panels **a**–**f** marks the average response time in each bin, *ms* milliseconds, *µV* microvolt. See text for further details

resulted because, in this time window, the overlap between a response-related positivity and the stimulus-related P300 was maximal which implies higher voltages according to the law of superposition. As response time increased above 600 ms (which occurred mainly with response-set size eight), a diagonal response-related positivity and a vertical stimulus-related positivity (P300) split up (indicated by an arrow in Fig. 2c). As a consequence, response-locked averages were more positive in the post-response period for trials with response times between 400 and 600 ms (which occurred mainly with response-set sizes two and four) than in trials with slower response times. This could explain why no clear Nc/CRN could be observed with response-set sizes two and four, and why no monotonous decrease of the Ne/ERN with increasing response-set size could be observed.

The distortions caused by the overlap of the stimulus-locked P300 with the response-locked components in the post-response window can be reduced by high-pass filtering (see, Luu and Tucker 2001; Vidal et al. 2003) which attenuates the contribution of the P300 to the overall ERP[4]. Figure 3 depicts the ERP distributions after application of a 4 Hz high-pass filter. Obviously, filtering strongly reduced the contribution of the stimulus-locked positivity to the post-response period between 400 and 600 ms. On correct trials, the Nc/CRN can now be seen as a response-locked negativity starting shortly before the response for all three response-set sizes (indicated by arrows on correct trials). On error trials, a clear Ne/ERN remains in the post-response window (indicated by arrows on error trials). These data were now re-analyzed using the same method to quantify Nc/CRN and Ne/ERN as above.

Base-to-peak analysis of average ERPs filtered with 4–10 Hz

Averages of the filtered data are depicted in Fig. 4. An Nc/CRN is now observable even in the two-choice task and the four-choice task, although it is still smaller in the two-choice task and four-choice task than in the eight-choice task (two, 2.83 µV; four, 1.88 µV; eight, 1.82 µV). In contrast, the Ne/ERN now decreases monotonically with increasing response-set size (two, 6.52 µV; four, 4.62 µV; eight, 4.20 µV). These observations were corroborated by

[4] An alternative method for removing slow potentials from EEG data is to calculate so-called Surface Laplacians (see, Vidal et al. 2003).

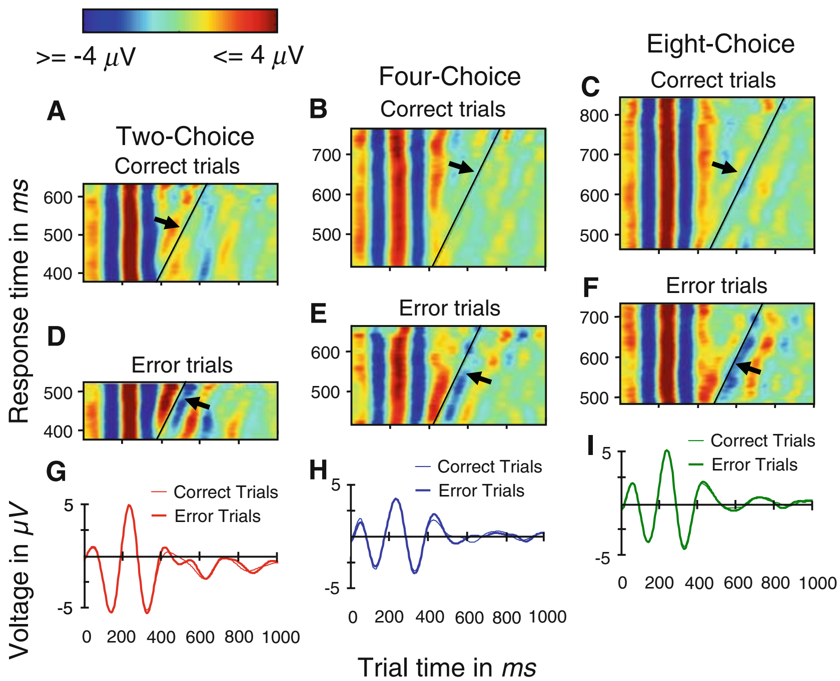

Fig. 3 Stimulus-locked voltage changes as a function of response time. ERPs were filtered by a 4 Hz high-pass filter and a 10 Hz low-pass filter, sorted by response time and assigned to overlapping response time bins of 20 ms width staggered by 5 ms. Averages for bins containing more than two trials were computed and plotted in a raster such that response time in milliseconds is represented on the ordinate, and trial time in milliseconds is represented on the abscissa. Positive voltages are depicted in *red*, and negative voltages are depicted in *blue*. Panels **a**, **b**, and **c** Correct trials in the two-choice task, the four-choice task, and the eight-choice task, respectively. Panels **d**, **e**, and **f** Error trials in the two-choice task, the four-choice task, and the eight-choice task, respectively. Panels **g**, **h**, and **i** Average stimulus-locked waveforms for correct trials (*thin lines*) and error trials (*thick lines*), respectively. The *diagonal black line* from the bottom to the top of panels **a** to **f** marks the average response time in each bin, *ms* milliseconds, *µV* microvolt. See text for further details

statistical analyses. An ANOVA comparing both components revealed a main effect of component indicating a larger Ne/ERN (5.08 µV) than Nc/CRN (2.18 µV), $F(1, 38) = 62.7$, $p < 0.001$, and, more importantly, an interaction of component and response-set size, $F(2, 38) = 6.83$, $p < 0.01$. Post hoc tests showed that the Ne/ERN was larger in the two-choice task than in the four-choice task, $t(25) = 1.96$, $p < 0.05$, and the eight-choice task, $t(25) = 2.40$, $p < 0.05$, whereas it still did not differ significantly between the latter two. The Nc/CRN was significantly larger in the eight-choice task than in the two-choice task, $t(25) = 1.68$, $p < 0.05$, and it was marginally significantly larger in the eight-choice task than in the four-choice task, $t(26) = 1.61$, $p = 0.06$, whereas it did not differ significantly between the four-choice task and two-choice task. Finally, the analysis of the difference wave peaks (see, Fig. 4, panel c) revealed a highly significant effect of response-set size (two, 5.97 µV; four, 3.67 µV; eight, 2.47 µV), $F(2, 38) = 9.82$, $p < 0.001$.

In sum, the post-response ERPs of both correct and error trials were clearly influenced by increasing response-set size. When response-set size was increased from two to four, the Ne/ERN on error trials clearly decreased. However, a further increase of response-set size to eight did not yield a significant further decrease of the Ne/ERN. The Nc/ERN on correct trials, by contrast, increased mainly when response-set size was increased from four to eight, whereas an increase from two to four did not yield a clear difference.

Discussion

Current theories on action monitoring predict that whatever impairs response selection also impairs the detection of errors. This results because error detection is assumed to depend on the detection of a mismatch or conflict between the error response and the correct response (e.g., Falkenstein et al. 1990; Yeung et al. 2004), and this, in turn, depends on the efficiency by which the correct response becomes represented or activated following the error. The goal of the present study was to test this prediction by considering the effects of response-set size on behavioral and electrophysiological correlates of action monitoring. Whereas response-set size has constantly been shown to affect response selection (e.g., Usher et al. 2002), its influence on action monitoring is less clear. Whereas an early study provided evidence that an increased response-set size impairs behavioral measures of error detection (Rabbitt

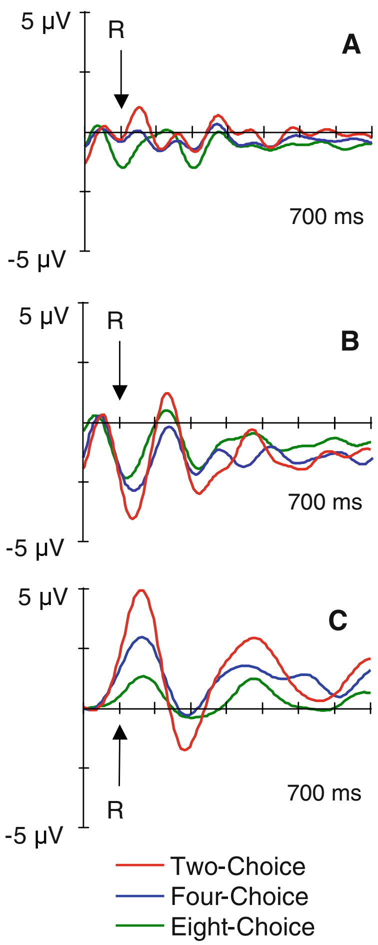

Fig. 4 Response-locked grand average waveforms filtered with a high-pass filter of 4 Hz and a low-pass filter of 10 Hz for correct trials (panel **a**), error trials (panel **b**), and correct trials minus error trials (panel **c**) in the two-choice task (*red lines*), in the four-choice task (*blue lines*), and in the eight-choice task (*green lines*), *ms* milliseconds, *μV* microvolt. See text for further details

1967), a recent study failed to find an effect of response-set size on the error-related ERP (Pailing and Segalowitz 2004). However, Pailing and Segalowitz (2004) used a rather small variation of response-set size by comparing a two-choice task and a three-choice task. Thus, in the present study, error-related brain activity was compared across participants performing a two-choice task, a four-choice task, or an eight-choice task.

As predicted, action monitoring was strongly affected by response-set size. Increasing response-set sizes implied decreased Ne/ERN amplitudes and increased Nc/CRN amplitudes. Moreover, it also led to an impaired post-error slowing as well as a trend toward longer error signaling latencies. However, clear effects of response-set size on post-response ERPs were only obtained when the contribution of the stimulus-locked P300 was attenuated by applying a 4 Hz high-pass filter. Without filtering, the stimulus-locked P300 distorted the post-response ERPs selectively for the two-choice and the four-choice tasks, leading to underestimation of Ne/ERN and Nc/CRN amplitudes in these conditions. This effect was weaker in the eight-choice task because the longer response times implied that the P300 preceded the response. This demonstrates the necessity of controlling for stimulus-locked potentials when comparing response-locked ERPs from conditions with different response times (see also, Coles et al. 2001). This effect could partially be responsible for the fact that Pailing and Segalowitz (2004) failed to observe an effect of response-set size on post-response ERPs.

As already mentioned, our results are in accordance with theories that attribute the Ne/ERN to processes involved in error detection. For instance, conflict monitoring theory (Yeung et al. 2004) assumes that the Ne/ERN represents the conflict between the correction response emerging after the error and the still active error response. Because an increased response-set size generally impairs response selection (Usher et al. 2002), it also impairs the correction response. This leads to a smaller amount of post-error conflict and thus to a smaller Ne/ERN. At the same time, it increases the risk that an erroneous response becomes activated after a correct response. This leads to post-response conflict on correct trials and thus to a larger Nc/CRN.

Similarly, the conflict monitoring account can also explain why behavioral indicators of error detection such as signaling latency and post-error slowing are impaired when response-set size is increased. According to this theory, error detection is achieved by detecting post-error conflict (Yeung et al. 2004). Therefore, a reduction of post-error conflict goes along with impaired error detection. The question arises, why the effect of response-set size on Nc/CRN was not accompanied by a corresponding increase in false alarms in error signaling. However, the rate of false alarms was very low (<1%) which might have prevented that an effect of response-set size was obtained. Presumably, the amount of post-response conflict on correct responses was large enough to produce an Nc/CRN but too small to trigger error detection.

A similar explanation can be derived from the mismatch hypothesis (e.g., Falkenstein et al. 2000) which assumes that the Ne/ERN represents a mismatch between the error response and the correct response. When the representation of the correct response is impaired because response-set size is increased, this should increase the risk of a miss on error trials as well as that of a false alarm on correct trials. Moreover, because the concept of a mismatch detector can also account for behavioral measures of error detection (Steinhauser et al. 2008), the fact that mismatch detection is impaired by increasing response-set size can also explain why this leads to impaired error signaling or post-error slowing.

However, both the conflict monitoring account and the mismatch hypothesis assume that the Ne/ERN reflects a signal that serves as the basis of error detection. Alternatively, one could assume that the Ne/ERN and the Nc/CRN reflects a later stage of error detection that is more related to the subjective uncertainty about the correctness of a response (see e.g., Scheffers and Coles 2000). For instance, Pailing and Segalowitz (2004) interpreted their results by assuming that stimulus discriminability but not response-set size influences this subjective uncertainty. Although the present results show that response-set size has an effect on the Ne/ERN and the Nc/CRN, this could simply imply that an increasing response-set size increases the subjective uncertainty about response correctness in the same way as stimulus discriminability, e.g., by increasing the perceived difficulty of a task. Therefore, further research is needed to decide which stage of error detection is reflected by these results.

Other accounts of the Ne/ERN relate the component not to post-response conflict or to a mismatch between the correct and the erroneous response, but to the significance of errors for ongoing behavior (e.g., Hajcak et al. 2005; Maier et al. 2008) or to behavioral adjustment (Holroyd and Coles 2002). For instance, the reinforcement learning theory of the Ne/ERN (Holroyd and Coles 2002) assumes the Ne/ERN to be a correlate of a reinforcement signal that is generated whenever an event is worse than expected. When a stimulus is presented and the correct stimulus–response mapping is known, the system expects a correct response to be produced. However, if an error occurs then this expectation is violated and behavioral adjustments are necessary to meet this expectation in the future. The reinforcement learning theory generally assumes that the Ne/ERN is negatively correlated with error probability across conditions. Accordingly, it could explain an effect of response-set size on the Ne/ERN provided that an increasing response-set size is associated with an increased error rate. However, because error rate is not affected by response-set size in our data, it is difficult to explain our results in terms of this theory[5].

[5] Although the error rate was the same in our three conditions, the absolute number of errors differed because of different trial numbers. To examine whether this can account for our results, we re-analyzed the data using only the first 1,280 trials from each condition. This did not change the results. Accordingly, our findings are not related to the different absolute numbers of errors in the three conditions. Another possibility is that our results reflect the different number of trials used for computing the waveforms. To control for this effect, a bootstrapping technique was applied. The smallest number of error trials in a single participant was 38. Therefore, we recomputed waveforms for each participant by randomly drawing 38 trials from each condition (correct responses and errors), and repeated this for 1000 times. For each repetition, components were quantified, and component measures were averaged across repetitions. The results obtained in this way were the same as in the main analysis, suggesting that different trial numbers were not responsible for our results.

Taken together, the results of the present study demonstrate that action monitoring suffers when response-set size is increased. An increased number of response alternatives implies a reduced Ne/ERN amplitude on error trials, an increased Nc/CRN amplitude on correct trials, as well as smaller post-error slowing and a trend toward less efficient error signaling. These results provide evidence for the general idea that whatever impairs response selection also impairs action monitoring. Moreover, they are in accord with theories that attribute the Ne/ERN to post-error response conflict (Yeung et al. 2004) or to a mismatch between the correct and the error response (Falkenstein et al. 1990, 2000). Finally, by showing that response-set size influences not only response times but also action monitoring, it demonstrates the relevance of the response-set for optimal and error-free performance. This represents an important constraint, e.g., for the construction of human–machine interfaces.

References

Alegria J, Bertelson P (1970) Time uncertainty, number of alternatives and particular signal–response pair as determinants of choice reaction time. Acta Psychol 33:36–44

Bernstein PS, Scheffers MK, Coles MG (1995) "Where did I go wrong?" A psychophysiological analysis of error detection. J Exp Psychol Hum Percept Perform 21:1312–1322

Carter CS, Braver TS, Barch DM, Botvinick M, Noll D, Cohen JD (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280:747–749

Coles MGH, Scheffers MK, Holroyd CB (2001) Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. Biol Psychol 56:173–189

Danielmeier C, Wessel JR, Steinhauser M, Ullsperger M (2009) Modulation of the error-related negativity by response conflict. Psychophysiology 46:1–11

Dehaene S, Posner MI, Tucker DM (1994) Localization of a neural system for error detection and compensation. Psychol Sci 5:303–305

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21

Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 34:1443–1449

Eriksen BA, Eriksen CW (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 16:143–149

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: Brunia CHM, Gaillard AWK, Kok A (eds) Psychophysiological brain research, 1. University Press, Tilburg, pp 192–195

Falkenstein M, Hoormann J, Christ S, Hohnsbein J (2000) ERP components on reaction errors and their functional significance: a tutorial. Biol Psychol 51:87–107

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

Gratton G, Coles MGH, Donchin E (1983) A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 55:468–484

Hajcak G, Vidal F, Simons RF (2004) Difficulties with easy tasks: ERN/Ne and stimulus component overlap. In: Ullsperger M, Falkenstein M (eds) Errors, conflicts, the brain. Current opinions on performance monitoring, 1. Max Planck Institute of Cognitive Neuroscience, Leipzig, pp 204–211

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

Hick WE (1952) On the rate of gain of information. Q J Exp Psychol 4:11–26

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

Laming D (1979) Choice reaction performance following an error. Acta Psychol 43:199–224

Logan GD (1988) Toward an instance theory of automatization. Psychol Rev 95:492–527

Luu P, Tucker DM (2001) Regulating action: alternating activation of midline frontal and motor cortical networks. Clin Neurophysiol 112:1295–1306

Luu P, Flaisch T, Tucker DM (2000) Medial frontal cortex in action monitoring. J Neurosci 20:464–469

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

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

Rabbitt PM (1966) Errors and error correction in choice-response tasks. J Exp Psychol 71:264–272

Rabbitt PMA (1967) Time to detect errors as a function of factors affecting choice-response time. Acta Psychol 27:131–142

Rabbitt PMA (2002) Consciousness is slower than you think. Q J Exp Psychol Hum Exp Psychol 55:1081–1092

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

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

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

Usher M, Olami Z, McClelland JL (2002) Hick's law in a stochastic race model with speed–accuracy tradeoff. J Math Psychol 46:704–715

Vidal F, Hasbroucq T, Grapperon J, Bonnet M (2000) Is the 'error negativity' specific to errors? Biol Psychol 51:109–128

Vidal F, Burle B, Bonnet M, Grapperon J, Hasbroucq T (2003) Error negativity on correct trials: a reexamination of available data. Biol Psychol 64:265–282

Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev 111:931–959