This is the accepted author manuscript, not the final typeset article. Please find the official publication at:

Foerster, A., Schaaf, M., Weissman, D. H., Kunde, W., & Pfister, R. (2026). Response activation in error processing: Assessing leakage into upcoming action episodes. J*ournal of Experimental Psychology: Human Perception and Performance*, *52*(4), 516–538. https://doi.org/10.1037/xhp0001402

Response activation in error processing:

Assessing leakage into upcoming action episodes

Anna Foerster[1], Moritz Schaaf[1*], Daniel H. Weissman[2], Wilfried Kunde[3] & Roland Pfister[1]

[1] Trier University, Johanniterufer 15, 54294 Trier, Germany

[2] University of Michigan, East Hall, 530 Church Street, Ann Arbor, MI 48109, USA

[3] University of Würzburg, Röntgenring 11, 97070 Würzburg, Germany

* corresponding author

Word count: 14231

Author Note

We have no conflicts of interest to disclose. This research has been funded through the project FO 1505/2-2, awarded to AF by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the Research Unit FOR2790, and through the Heisenberg Programme of the DFG, awarded to RP (PF 853/10-1, PF 853/11-1). We preregistered the experiments that we conducted (Experiment 1: osf.io/m2v5a; Experiment 2: osf.io/f49z7) and made the program code, data, and analysis code publicly available (osf.io/f49z7; Foerster, Schaaf, et al., 2022).

Abstract

Error commission is accompanied by a cascade of cognitive processes. In current views, these processes change the activation of all task-related responses in the same way (e.g., via generalized inhibition or a shift toward a more conservative response criterion). Using a three-choice task, we contrasted such response-general processes with putative response-specific processes, which are not sufficiently incorporated into contemporary views. We hypothesized that response-specific processes influence both the correct response and the erroneous response but not a third (i.e., neutral) response. To test this hypothesis, we assessed the activation of these three responses throughout the task using finger force and/or standard keypresses. Replicating prior findings, we observed evidence that response-specific processes attempt to cancel and correct the execution of hastily initiated errors. Crucially, analyses of performance in post-error trials showed that both the correct response and the erroneous response from the previous error trial were more active than the neutral response. These novel findings suggest that response-specific processes explain critical aspects of post-error performance.

Keywords: Error processing, error cancellation, error correction, inhibition, conflict

Public significance statement

Errors are accompanied by cognitive processes that shape ongoing and future actions. Contemporary views posit that such processes affect all task-related responses in the same way. We show for the first time that some of these processes heighten the activation of erroneous and correct responses more than they heighten the activation of neutral responses. These novel findings suggest that response-specific cognitive processes contribute to post-error performance.

Introduction

Humans err. When this happens, numerous cognitive processes are set into motion that not only mitigate ongoing error commission but also change the course of subsequent action. The current study seeks to distinguish between *response-general* and *response-specific* error processing. Most prominent theories of error processing implicitly assume that only response-general processes (i.e., processes that exert an equivalent impact on *all* possible behavioral responses) influence post-error performance. In the present study, however, we challenge this implicit assumption by investigating whether response-specific processes (i.e., processes that exert a selective influence on *individual* responses) also make an important contribution.

The commission of an error

While errors can arise from various sources like erroneously applying a wrong task-set (e.g., Moretti et al., 2021, 2023), most theories of error processing focus on errors that occur when participants adopt the correct task-set. Such commission errors, or action slips (Reason, 1990, 1995), occur when an executed action does not fit with the current action goal (e.g., hitting a switch that rings a doorbell in the hallway when the goal is to turn on the light by hitting the adjacent switch). To investigate commission errors, researchers typically ask participants to follow rules about how to respond to a set of target stimuli with specific response keys. For example, they might ask participants to respond to the letter *X* with a left key, the letter *Y* with a middle key, and the letter *Z* with a right key. Even for such simple rules, commission errors happen from time to time, like responding to the letter X by pressing the middle key.

Commission errors have behavioral signatures beyond the fact that the action is erroneous. Errors are often initiated more quickly, suggesting that they may occur particularly frequently when pre-response planning is insufficient (e.g., when using a response threshold that is too low; e.g., Brewer & Smith, 1989; Rabbitt, 1966, 1979). Furthermore, commission errors are typically executed with less force and terminated earlier than correct responses (e.g., Bode & Stahl, 2014; Foerster, Steinhauser, et al., 2022; Hochman et al., 2017; Pfister et al., 2025), which suggests that the cognitive system compensates for poor planning by immediately cancelling erroneous actions.

Interestingly, participants are especially quick to deliver a correction response after erring (e.g., Fiehler et al., 2004, 2005; Rabbitt, 1966, 2002). That is, the average time between committing an error and making a corrective action is substantially shorter than the average time between stimulus

presentation and a correct response. These observations suggest that the correct response is generated independently of – and in parallel with – the erroneous response via the continued processing of the target. Put another way, these observations suggest that the cancellation of the erroneous response together with the activation of the correct response are *response-specific* error processes.

The coactivation of the correct and the erroneous response is a central assumption of the conflict monitoring model (Botvinick et al., 2001; Yeung et al., 2004). This recurrent neural network model proposes a constantly active monitoring process that detects conflict arising from concurrently active cognitive representations. In the case of errors, there is conflict between the representation of the executed erroneous response and the representation of the concurrently prepared correct response. Therefore, in this model, error detection is not special as it appears to rely on the same monitoring process that identifies other types of conflicts (e.g., competing response tendencies induced by incongruent stimuli even when a correct response occurs). According to conflict monitoring theory, the detection of any conflict increases the recruitment of cognitive control, which counteracts conflicting response activation and biases subsequent actions towards a more conservative response criterion (Botvinick et al., 2001).

The aftermath of erring

The most prominent behavioral change after committing an error in trial n is the slowing of responses in the subsequent trial n+1 (e.g., Laming, 1979; Rabbitt, 1966). Explanations of such post-error slowing posit that errors induce a shift toward a more conservative response criterion (e.g., Botvinick et al., 2001; Dutilh et al., 2012; Jentzsch & Leuthold, 2006), lead participants to orient attention away from the task (e.g., Houtman et al., 2012; Notebaert et al., 2009; Notebaert & Verguts, 2011; Núñez Castellar et al., 2010; Steinborn et al., 2012), or trigger intensive monitoring that occupies information processing (e.g., Houtman & Notebaert, 2013; Jentzsch & Dudschig, 2009; Schaaf et al., 2022; Welford, 1959). Crucially, these suggested processes can be subsumed under our definition of *response-general* processes because they apply to *all* responses in a task.

Here, we suggest that response-specific processes might also influence post-error behavior. Indeed, two recent findings suggest that representations of the erroneously executed response and the not executed correct response from trial n are especially active in trial n+1. First, the erroneous response from trial n is initiated more quickly in trial n+1 than a neutral response alternative from trial n that was neither the correct response nor the erroneous response (Foerster et al., 2023). Second, during an error

trial, bindings (i.e., associations) form between (i) the acted-upon stimuli and the not executed correct response and (ii) the executed erroneous response and its perceptual consequences (Foerster, Moeller, et al., 2022; Foerster et al., 2021; Parmar et al., 2022). Retrieving such bindings might make it easier to initiate either the correct or the erroneous response from trial n than to initiate the neutral response.

Crucially, these recent findings cannot be explained by models that posit only response-general processes following an error. First, the generalized inhibition model assumes that cognitive inhibition decreases the activation of all current representations, such as the active task representation, to enable a switch to other representations, such as the source of the error (Wessel, 2018). Applying inhibition in such a general manner, however, should render the correct, erroneous, and neutral response from the erroneous trial n equally less accessible in trial n+1. Second, computational models of conflict monitoring assume quick, mutual inhibition of active response representations (Yeung et al., 2004), and post-error behavioral changes are typically explained with a shift in the tradeoff between speed and accuracy (Botvinick et al., 2001). Critically, adopting a more conservative response criterion for all responses is another example of a response-general process, rather than a response-specific process that can differentially alter the activation of specific responses.

The present study

Given that contemporary views of error processing posit that only response-general processes contribute to post-error performance, we investigate whether and when response-specific processes[1] also contribute. Specifically, we investigate error-related processes that are (i) *response-specific* and operate *during* error commission (i.e., error cancellation and correction), (ii) *response-specific* and operate *after* an error (i.e., activation or inhibition of the correct and/or erroneous response), and (iii) *response-general* and operate *after* an error (i.e., generalized inhibition or conservative criterion shifts).

The most critical contribution of this research is the investigation of response-specific processes after an error. Therefore, we employed a speeded three-choice task and hypothesized that response-specific processes would influence the correct response and the erroneous response but not the third (i.e., neutral) response. To test this hypothesis, we assessed the activation of these three responses by measuring finger force (Experiment 1 only) and behavioral performance (Experiment 1 and 2) following both errors and correct trials.

[1] When referring to a post-error process, we do not intend to imply that the process is necessarily active or unique to error trials. For instance, we would consider residual activation of a particular finger to be a response-specific process.

The present study adds to the literature in several ways. First, it extends a previous experiment that investigated response-specific activation of the executed erroneous response in the trial after an error (Foerster et al., 2023). Second, it provides novel evidence for activation of the not executed correct response in the trial after an error. Third, it disentangles for the first time the effects of such response-specific activation on performance from the effects of response-general processes that have been investigated extensively without considering response-specific processes (e.g., Dutilh et al., 2012; Jentzsch & Dudschig, 2009; Notebaert et al., 2009). Finally, we note that although previous research provides ample evidence for response-specific processes that operate *during* commission errors (e.g., Bode & Stahl, 2014; Foerster, Steinhauser, et al., 2022; Hochman et al., 2017), we investigate these processes again because error cancellation and error correction motivate hypotheses about response-specific processing *after* error commission (e.g., about inhibiting or activating specific responses).

Experiment 1

Introduction

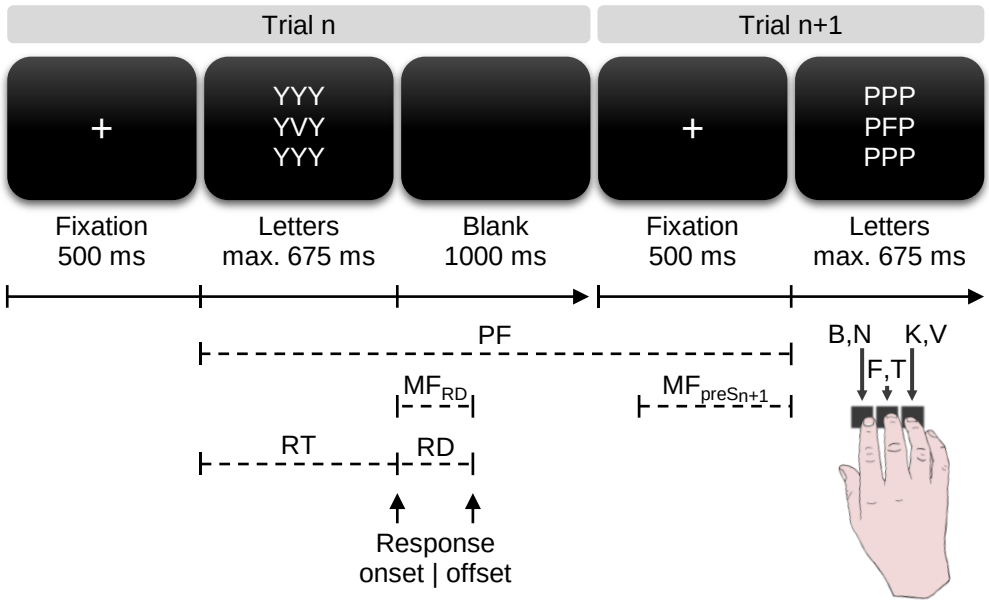
To assess response-specific versus response-general error-related processes, Experiment 1 employs a speeded forced-choice task with three response options (see Fig. 1). This task allows us to disentangle the influences of (i) the executed, erroneous response from (ii) the not executed, correct response and (iii) the not executed, neutral (i.e., third) response that corresponds to neither the erroneous response nor the correct response. More specifically, this task allows us to investigate whether behavioral changes after an error apply generally to all responses or only to specific responses. Response-general processes should affect the neutral response to the same extent as the erroneous and correct response. Response-specific processes, however, should affect the erroneous and/or correct response differently than the neutral response.

Since participants respond using force-sensitive keys, we can also examine subtle changes in motor activity with high temporal resolution. Therefore, we can supplement the insights from overt responses with data about sub-threshold motor activation during error commission and during seemingly idle times; that is, during the response-stimulus interval. Since sub-threshold response force has already provided valuable insights into the cognitive underpinnings of correct actions in the response-stimulus interval (Weissman, 2019; Weissman & Schmidt, 2024; Weissman et al., 2024), we here explore its potential for tracking how error-related processes unfold over time.

Because the data are rich and unfold across time, several analyses are required to capture all aspects of interest. In a nutshell, we investigated whether error commission shows signs of poor pre-response planning (Hypothesis 1, H1), and whether it is characterized by response-specific cancellation (H2) and correction (H3). We further scrutinized whether evidence for response-specific activation or inhibition emerges after an error in sub-threshold responses (H4) and overt responses (H6). At the same time, we tested for response-general inhibition or conservatism in sub-threshold responses (H5) and overt responses (H7). To improve clarity, we specify our hypotheses in the *Results* section prior to indicating the corresponding empirical results (see Table 1).

Figure 1

Trial structure and dependent measures in Experiment 1



Note. After a fixation of 500 ms, participants respond to a central target letter. To provoke commission errors, eight neutral, irrelevant letter stimuli surround the central target letter, and a short response deadline of 675 ms requires quick responses. Upon detecting the onset of a response, the screen goes blank for 1000 ms before the next trial (i.e., trial n+1) starts. The six possible letter targets are mapped to three force-sensitive response keys. Response time (RT) is the time from letter onset to response onset, and response duration (RD) is the time from response onset to response offset of the above-threshold response. Peak force (PF) indicates maximal force of sub-threshold responses during the stimulus-stimulus interval (i.e., the time between the stimulus onsets in trial n and trial n+1). MF_{RD} denotes the mean force of the above-threshold response between response onset and offset. $MF_{PreS_{n+1}}$ represents the mean force of sub-threshold responses during a pre-stimulus time interval (i.e., 400 ms before the onset of the letter stimuli in trial n+1).

Method

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. We preregistered this study before data collection (osf.io/m2v5a), which occurred in 2022 and 2023. The E-Prime code that we used to program the experiment, data, and analysis code are publicly available (osf.io/f49z7). We processed and analyzed the data in *R* version 4.3.1 (R Core Team, 2023), using the packages *schoRsch* (Pfister & Janczyk, 2016), *data.table* (Barrett et al., 2024), *tidyverse* (Wickham et al., 2019), *mousetRajectory* (Pfister et al., 2024), *ez* (Lawrence, 2016), and *patchwork* (Pedersen, 2023).

Participants

A reanalysis of published data with only two response options suggested large effects for differences between correct and erroneous responses in response time (RT; d_z = 0.91), response duration (RD; d_z = 1.80), and mean force (MF; d_z = 1.21; Foerster, Steinhauser, et al., 2022). At an alpha of .05, 15 participants would provide a power of > .95 to detect these effects (calculated with the *power.t.test* function in *R* version 4.0.3; R Core Team, 2023). Predictions for the effect sizes of the other hypotheses could, however, not be obtained from the existing literature. Thus, we preregistered to collect 48 analyzable datasets (i.e., replacing excluded participants), thereby ensuring a power of .95 to detect effects of $d_z \geq 0.5$ for both directional and nondirectional tests. Further, we preregistered to analyze the first eight participants, continuing data collection of the remaining sample only if at most two datasets (≤ 25%) had to be discarded based on our exclusion criteria. While exclusion rates in four pilot samples were too high (see *Appendix A*), the data from the first eight participants in the version of the task that we report in Experiment 1 (i.e., the fifth pilot sample) allowed us to continue data collection.

Forty-seven participants who provided written informed consent and received monetary compensation were included in the final sample (mean age = 27, *SD* = 8.7; 38 female, nine male, none non-binary; 41 right-handed, five left-handed, one ambidextrous). Data from 11 additional participants were collected but excluded from all analyses. As preregistered, seven participants were not invited to the second session because they responded correctly in less than 50% of the trials in the first session. Accidentally, however, two additional participants that responded correctly in less than 50% of the trials during the first session were invited to the second session. In the initial analysis of the data for determining exclusions and replacements, we did not differentiate between accuracy in the first and second session by mistake. Only one of these two participants was replaced based on this accuracy criterion. The other participant was excluded but could not be replaced because this error went unnoticed until the experimental setup had already been disassembled. For one additional participant, the experimental program crashed in the second half of the second session. Although we had not anticipated this scenario in the preregistration, we had to exclude and replace this dataset. Finally, and as preregistered, one participant was excluded and replaced because they had less than 10 observations in a design cell for the main analyses (see *Data treatment*).

The study was performed in accordance with the ethical guidelines of the German Psychological Society (DGPs) and with the Declaration of Helsinki. Our funding agency, the DFG, as well as the ethics committee of the University of Würzburg where the study was conducted, specify under which

circumstances approval for a psychological study is required. In particular, approval is required if there is a risk or a high level of stress for participants, information about the aims and procedures are only partially provided, patients participate in the study, or if a study involves functional magnetic resonance imaging, electrical or magnetic stimulation, or psychopharmacological treatments. As none of these conditions applied to our study, we did not seek ethical approval by an institutional review board.

Stimuli and apparatus

The stimuli consisted of letters presented in white font against a black background, displayed on a 24″ screen. A relevant target letter (one of the letters *B, F, K, N, T, V*) was surrounded by eight identical, neutral, irrelevant letters (one of the letters *C, H, M, P, S, Y*), thus forming a 3 × 3 letter grid (see Fig. 1). Only the six targets were mapped to three custom-built keys, which participants pressed with their right index, middle, and ring fingers. We used irrelevant letters that did not map onto responses but introduced visual noise to potentially provoke error commission (Foerster, Moeller, et al., 2022). We counterbalanced the assignment of letter pairs to response keys (e.g., *B* and *N* to the left key; *F* and *T* to the middle key; *K* and *V* to the right key) across participants. The stimulus-response mapping remained constant within each participant across both experimental sessions.

The custom-built keys measured the isometric force applied to them with capacitive sensors. They did not provide tactile feedback on whether they were pressed. The participants' fingers were not fixed to the keys, but we instructed participants to keep their fingers on the keys during the whole block. Increasing force values indicate stronger downwards movement of a finger while decreasing force values indicate decreasing downwards movement. In contrast to previous research using trial-wise baseline measurements (Foerster, Steinhauser, et al., 2022), the baseline of the current experiment was the key-specific average force within a 100 ms window at the beginning of each block. To ensure that this was an accurate baseline, we explicitly instructed subjects to put their index, middle, and ring finger on the keys before the self-paced start of each block. However, participants could completely lift their fingers from the keys if they wanted to. We implemented this baseline measure at the beginning of each block because we explicitly investigated sub-threshold force in the interval that previously served as the baseline measurement interval. In line with previous research (Foerster, Steinhauser, et al., 2022), response on- and offsets were computed relative to an individual threshold, and this threshold was defined as the baseline force plus 225 arbitrary units (a.u.; approximately 2.5 N or 250 g). Response onset was defined as the first instance where the current force measurement (sampled at 250 Hz) and

four consecutive measurements were at or above threshold. Response offset was defined as the first instance thereafter where the force measurement fell below the threshold and remained below the threshold for another four consecutive force measurements.

Procedure

The experiment consisted of two sessions separated by one to seven days. Due to illness, one participant completed their second session only after 14 days. Both sessions started with a practice block of 30 trials, followed by 15 experimental blocks of 60 trials each (930 trials per session, 1860 in total). Each trial started with the onset of a central fixation cross. After 500 ms, the target letter and irrelevant letters appeared and remained visible until either the onset of a response was detected or the response deadline of 675 ms elapsed. During the practice blocks, feedback was displayed for 2000 ms after every trial. The feedback was “Good!” (German: “Gut!”) in green font for correct responses, “Too early!” (German: “Zu früh!”) in red font for premature responses during fixation, “Wrong!” (German: “Falsch!”) in red font for incorrect responses after target onset, and “Too slow” (German: “Zu langsam”) in red font for omission errors. During experimental blocks, feedback was only displayed for premature responses or omission errors. The next trial started after a blank inter-trial interval of 1000 ms.

To minimize potential influences of stimulus repetitions, neither the target nor the irrelevant letter was repeated in consecutive trials. Target letters were randomized across an experimental block. However, when a target was about to repeat in two consecutive trials, it was replaced with the other target letter that required the same response. This replacement ensured that participants could not anticipate the upcoming response identity even though they could potentially derive expectations about the upcoming target letter. To prevent repetitions of irrelevant letters, blocks were divided into sequences of length six. For these sequences, the set of six irrelevant letters was randomized under the constraint that the first irrelevant letter of a sequence always differed from the last two irrelevant letters of the preceding sequence. After each block, participants received feedback regarding their performance and had the opportunity for a self-paced break. Before beginning the next block, they were encouraged to respond as fast as possible without committing too many errors and were reminded to keep their fingers on the keys during the entire block.

Data treatment

For three participants, the experiment had to be re-started in the first session due to technical issues. We excluded the trials that were collected prior to the restart, although we did not anticipate and specify this in the preregistration. We also excluded spurious measurements from the unprocessed force data as occasionally all three force keys provided a value of -10 Volts / 0 a.u. at the same time (1.3%) due to internal calibration of the analog-to-digital converter. All other data treatment went as planned in the preregistration.

We extracted RTs, RDs, and peak forces (PFs) before further processing the force data. RT denotes the time between letter onset and response onset (see Fig. 1). RD denotes the time between response onset and response offset (Pfister et al., 2023; see also e.g., Foerster, Steinhauser et al., 2022; Pfister et al., 2022; Tonn et al., 2023). RT and RD were assessed for the above-threshold (i.e., executed) response. Finally, PF denotes the maximal force on a key during a time interval that starts with the onset of the target in trial n and ends with the onset of the target in trial n+1 (i.e., the stimulus-stimulus interval). Since PF was intended to capture sub-threshold activation of the correct response (i.e., correction efforts) in erroneous trials, PF was only computed for erroneous trials and for keys for which the force did not exceed the response threshold. Further, a rather broad time interval was selected to account for temporal variability of correction efforts.

We resampled the force data (originally sampled effectively every 4.3 ms, *SD* = 0.07 ms) to 1000 Hz with linear interpolation. The interpolated data compensated for CPU-load dependent variations in sampling rate and provided a force measurement every millisecond. Resampling therefore allowed for averaging force across trials for visualization (see Fig. 2 for a visualization of the stimulus-locked force distributions as a function of response option identity in correct and erroneous trials) and capturing MF exactly for the targeted time window.

Next, we extracted the response-specific MF (i) between response onset and response offset in trial n (MF_{RD}), (ii) during the 400 ms idle time prior to the onset of the target stimulus (i.e., the pre-stimulus interval) in trial n+1 ($MF_{PreS_{n+1}}$), and (iii) between response onset and response offset in trial n+1 ($MF_{RD_{n+1}}$; see Fig. 1). MF_{RD} and $MF_{RD_{n+1}}$ were computed only for the above-threshold (i.e., executed) response. For $MF_{PreS_{n+1}}$, the specific pre-stimulus interval was selected to capture sub-threshold response activation without spurious biomechanical influences from the execution of the previous

response; that is, sufficiently long after response execution in trial n (Foerster, Steinhauser, et al., 2022; Weissman, 2019).

We excluded several types of trials. First, we excluded the trials from the practice block in each session. Second, we excluded the first and last trial of each experimental block as these trials had no preceding or ensuing trial. Third, we excluded trials with an error in the preceding trial n-1 (18.9%) to control for the potential influence of an immediately preceding error.

To compare performance for commission errors and correct responses and to determine their impact on subsequent performance, we selected only trial sequences with a correct response or a commission error in trial n and in trial n+1. Therefore, we excluded trial sequences with early responses during fixation (< 0.1%), omission errors (7.4%), and additional responses during the blank screen[2] (2.7%) in trial n. We further excluded trial sequences with early responses during fixation (< 0.1%), omission errors (7.1%), and additional responses during the blank screen (2.5%) in trial n+1. For this selection of trial sequences, we computed and analyzed the percentage of errors in trial n+1 (PE_{n+1}; the number of commission errors divided by the sum of commission errors plus correct responses).

In our analyses of all other dependent measures (i.e., all measures except PE_{n+1}), we made the following additional exclusions. First, we excluded trial sequences with commission errors in trial n+1 (6.6%). Second, we excluded outliers separately for each participant, dependent measure (RT, RD, MF_{RD}, PF, $MF_{PreS_{n+1}}$, RT_{n+1}, RD_{n+1}, and $MF_{RD_{n+1}}$), and corresponding design cell of the planned analyses. A trial-specific value was considered an outlier if the median distance of that value from every other value within the same participant, dependent measure, and design cell was 3 times larger than the median absolute distance of every value from every other value. Put differently, a value (e.g., an RT) was considered an outlier if it was greater than 3×Sn (Rousseeuw & Croux, 1993; code adapted from Jones, 2019). We excluded a trial if any dependent measure was considered an outlier (13.8%).

After these exclusions, one participant had less than 10 observations in a design cell for the analyses related to H1-H5 and, therefore, was excluded and replaced (see *Participants*). For H6, H7, and the planned exploratory analyses, we assessed whether the relevant cells contained at least 5

[2] While these additional responses include overt error correction responses, we were specifically interested in whether there is evidence for *sub-threshold* error correction efforts. Therefore, the current study controlled for successful correction activities by excluding such trials. However, a comparison of response-specific and response-general processes between sub-threshold correction efforts and overt corrections could be a fruitful avenue for future research.

observations. Two participants did not fulfill this criterion and were excluded without replacement from these analyses (but not from the analyses related to H1 through H5).

Results

In the following, we differentiate the analyses with respect to whether they focus on processes that occur *during* error commission (H1-H3), *after* error commission during the pre-stimulus interval of trial n+1 (H4 and H5), or *after* error commission during the response in trial n+1 (H6 and H7). H6 and H7 were planned but marked as exploratory in our preregistration because we accepted less observations per cell (see *After error commission in trial n+1: Responding*). Unplanned, exploratory analyses are reported after the discussion of the planned analyses.

Table 1 provides a summary of the assumed processes for the different time points and hypotheses, hypothesized effects, and empirical findings. The two central factors in all our analyses are (i) *response accuracy in trial n* and (ii) *response option identity in trial n*. For erroneous trials, response option identity in trial n distinguishes between the executed erroneous response, the not executed correct response, and the not executed neutral response (which is neither the executed erroneous response nor the not executed correct response). For correct trials, response option identity distinguishes between the executed correct response and the average of the two not executed neutral responses. For brevity, we omit the adjectives “executed” and “not executed” when reporting the results.

Table 1

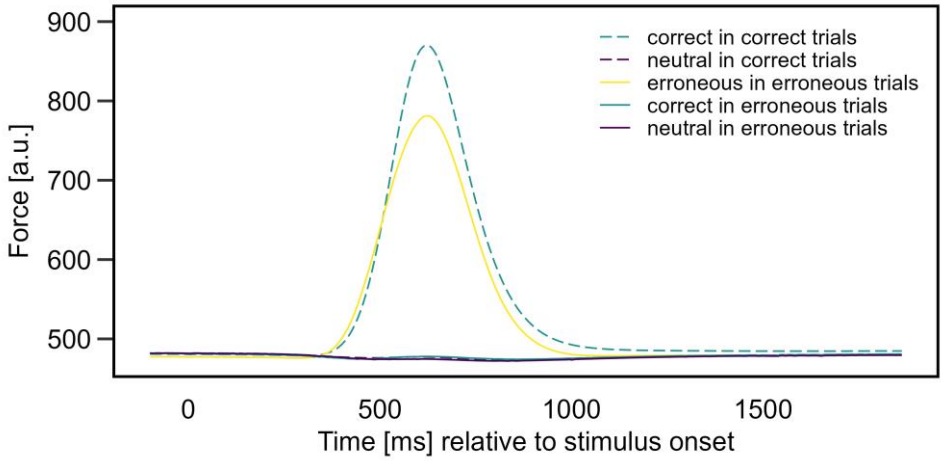
Timing, processes, hypotheses, and outcome of Experiment 1

Error process	H	DV	Hypothesized effect	Outcome
		<u>During error comission in trial n</u>		
Poor pre-response planning	1	RT	erroneous trial n < correct trial n	✔
Response-specific cancellation	2	RD	erroneous trial n < correct trial n	✔
		MF_{RD}	erroneous trial n < correct trial n	✔
Response-specific correction	3	PF	correct response > neutral response from erroneous trial n	✔
		<u>After error commission in trial n+1: Pre-stimulus interval</u>		
Response-specific inhibition	4	$MF_{PreS_{n+1}}$	erroneous response < correct/neutral response from erroneous trial n	
Response-specific activation	4	$MF_{PreS_{n+1}}$	correct/erroneous response > neutral response from erroneous trial n	
Response-general inhibition or conservatism	5	$MF_{PreS_{n+1}}$	neutral response from erroneous trial n < from correct trial n	
		<u>After error commission in trial n+1: Responding</u>		
Response-specific inhibition	6	RT_{n+1}	erroneous response > correct/neutral response from erroneous trial n	
Response-specific activation	6	RT_{n+1}	correct/erroneous response < neutral response from erroneous trial n	✔
Response-specific inhibition	6	PE_{n+1}	erroneous response < correct/neutral response from erroneous trial n	
Response-specific activation	6	PE_{n+1}	correct/erroneous response > neutral response from erroneous trial n	
Response-general inhibition or conservatism	7	RT_{n+1}	neutral response from erroneous trial n > from correct trial n	✔
	7	PE_{n+1}	neutral response from erroneous trial n < from correct trial n	

Note. Overview of hypotheses (H) 1-7 to test error processes assumed to operate during error commission, after error commission in the pre-stimulus interval, or after error commisson during responding. The table further specifies the hypothesized effects for each dependent variable (DV). A check mark indicates that the data produced an outcome that is consistent with the hypothesized effect.

Figure 2

Force profile in Experiment 1



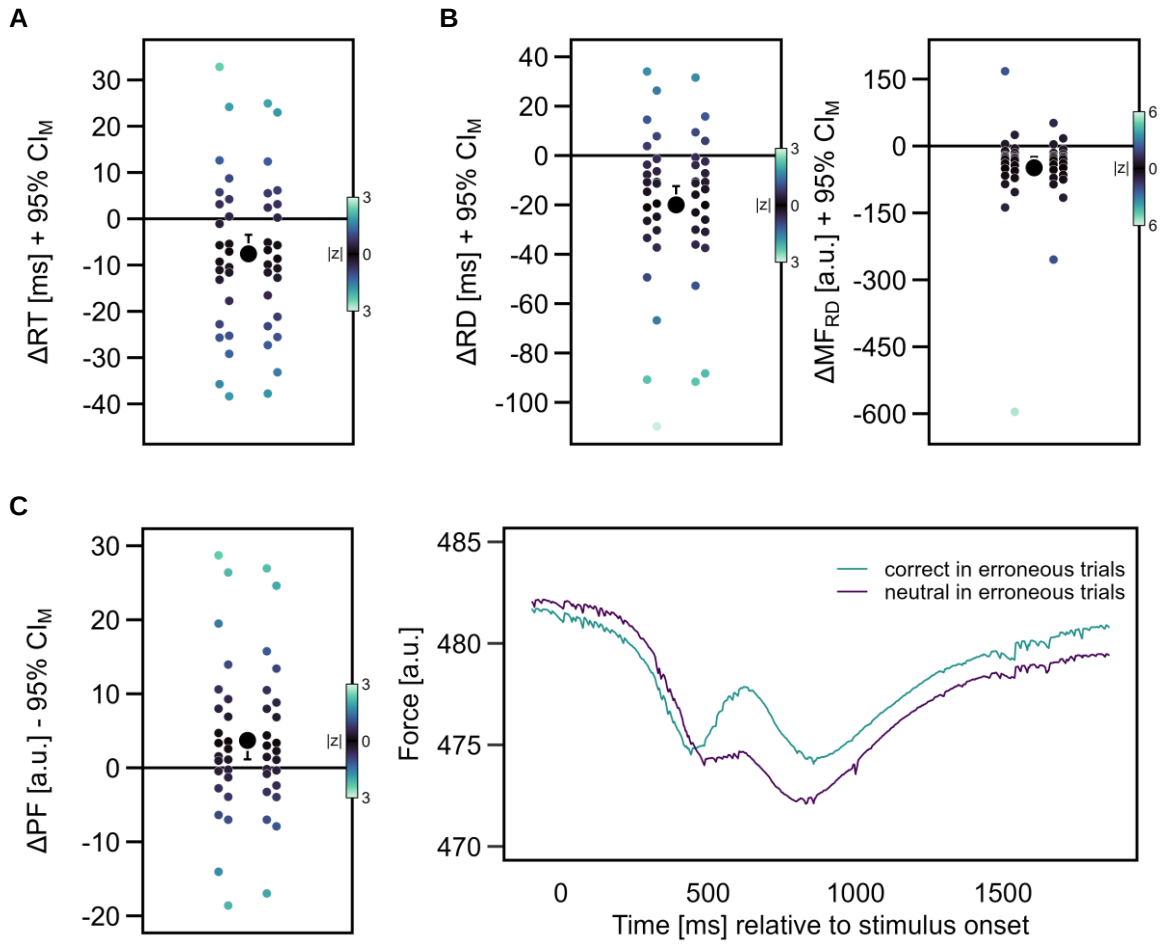
Note. Force distribution in arbitrary units (a.u.) locked to stimulus onset. Correct trials are indicated by dashed lines and erroneous trials are indicated by solid lines. Teal lines depict force on the correct response key (only executed in correct trials). Violet lines depict force on the neutral response keys (averaged across both neutral responses in correct trials). The yellow line depicts force on the erroneous response key in erroneous trials. Values reflect raw force without baseline-correction.

During error commission in trial n

In H1 to H3, we investigated processes during error commission. Specifically, H1 targeted poor pre-planning of errors, H2 targeted the response-specific process of error cancellation during error commission, and H3 targeted the response-specific process of error correction during the stimulus-stimulus interval (see Fig. 3).

Figure 3

Results for H1-H3 in Experiment 1



Note. Large, black dots indicate mean differences. Error bars visualize a one-tailed *t*-test. They correspond to the upper or lower bound of the 95% confidence interval of the related difference between means (CI_M). Small dots indicate individual participant values, and color visualizes the standardized deviation from the mean difference (|z|). Differences (Δ ≔ erroneous trial n – correct trial n) in **(A)** response time (RT; H1) and **(B)** response duration (RD; H2) and mean force of the response (MF_{RD}; H2). **(C)** Left: Differences (Δ ≔ correct response – neutral response) in peak force (PF; H3) in erroneous trials n. Right: Sub-threshold force distribution in arbitrary units (a.u.) locked to stimulus onset in erroneous trials n. Values reflect raw force without baseline-correction.

H1 predicts that erroneous responses will be initiated faster than correct responses due to less pre-response planning. Consistent with this hypothesis, a one-tailed paired-samples *t*-test revealed significantly faster RT in erroneous trials than in correct trials (519 vs. 527 ms; see Fig. 3A), $t(46) = -3.09$, $p = .002$, $d_z = -0.45$.

H2 predicts that the execution of an erroneous response will be shorter and less forceful than the execution of a correct response due to weaker initiation and active cancellation of erroneous responses. Consistent with this hypothesis, separate one-tailed paired-samples *t*-tests revealed significantly shorter

RD and lower MF_{RD} in erroneous trials than in correct trials (see Fig. 3B; RD: 159 vs. 179 ms, $t(46) = -4.42$, $p < .001$, $d_z = -0.64$; MF_{RD}: 839 vs. 887 a.u., $t(46) = -3.38$, $p = .001$, $d_z = -0.49$).

H3 predicts that correction efforts during error commission will result in higher sub-threshold force of the correct response compared to the neutral response. Consistent with this hypothesis, a one-tailed paired-samples *t*-test revealed significantly greater PF for the correct response than for the neutral response during erroneous trials (496 vs. 492 a.u.; see Fig. 3C), $t(46) = 2.43$, $p = .010$, $d_z = 0.35$.

In *Appendix B*, we explore response competition from correction efforts as an alternative interpretation for the observed differences between erroneous and correct responses during response execution. The exploratory analyses did not support this interpretation. In sum, the results are consistent with error cancellation and error correction efforts as response-specific processes.

After error commission in trial n+1: Overview

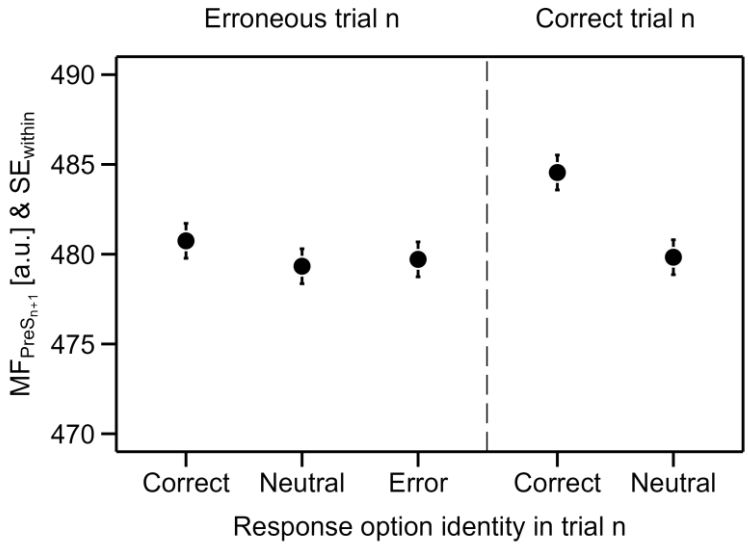
The following hypotheses concern the relative contributions of response-specific versus response-general processes after error commission in trial n. In our descriptions of these hypotheses, we refer to *response option identity in trial n*; that is, to the correct response, the erroneous response, and the neutral response. First, response-specific processes should influence the correct and erroneous response differently than the neutral response. Thus, they should produce differences in trial n+1 performance for these three responses following an error in trial n. Comparing trial n+1 performance for the correct response or the erroneous response to the neutral response should reveal whether response-specific processes inhibit or activate the correct and erroneous responses, respectively. Second, response-general processes should influence all three responses in the same way. To test this hypothesis, we determined whether committing an error versus responding correctly in trial n influences trial n+1 performance for the neutral response. This comparison offers the purest index of response-general processing changes because, unlike the correct and erroneous responses, the neutral response is not affected by response-specific processes. Third, and finally, we note that both response-specific and response-general processes could influence subthreshold force during the pre-stimulus interval and/or the overt response in trial n+1. We address these two possibilities below.

After error commission: The pre-stimulus interval of trial n+1

In H4 and H5, we investigated sub-threshold force changes that manifest during the pre-stimulus interval of trial n+1 following an error in trial n. Specifically, H4 targeted response-specific activation or inhibition, whereas H5 targeted response-general inhibition or conservatism (see Fig. 4).

Figure 4

Results for H4 and H5 in Experiment 1



Note. Mean sub-threshold force during the pre-stimulus interval in trial n+1 ($MF_{PreS_{n+1}}$). The left side of the figure shows the correct, neutral, and erroneous response from an erroneous trial n. The right side of the figure shows the correct and neutral response from a correct trial n. The error bars depict the within-subject standard error for the full factorial design (SE_{within}; Loftus & Mason, 1994).

H4 predicts that $MF_{PreS_{n+1}}$ will differ between the three response options after error commission in trial n, indicating response-specific changes. Inhibition of the erroneous response after its cancellation should lead to lower $MF_{PreS_{n+1}}$ for the erroneous response than for the correct or neutral response from trial n. In contrast, activation of relevant responses after error commission should lead to higher $MF_{PreS_{n+1}}$ for the correct and/or erroneous response than for the neutral response from trial n. To investigate these predictions, we analyzed $MF_{PreS_{n+1}}$ after erroneous trials using an ANOVA with a single within-subjects factor: Response option identity in trial n (correct vs. neutral vs. erroneous). This main effect was not significant, $F < 1$, which is inconsistent with H4.

H5 predicts that $MF_{PreS_{n+1}}$ will be lower for the neutral response after an erroneous trial than after a correct trial, indicating response-general changes such as inhibition or a shift to a more conservative

response threshold after committing an error. Inconsistent with H5, a one-tailed, paired-samples *t*-test comparing $MF_{PreS_{n+1}}$ in the two conditions above was not significant, $|t| < 1$.

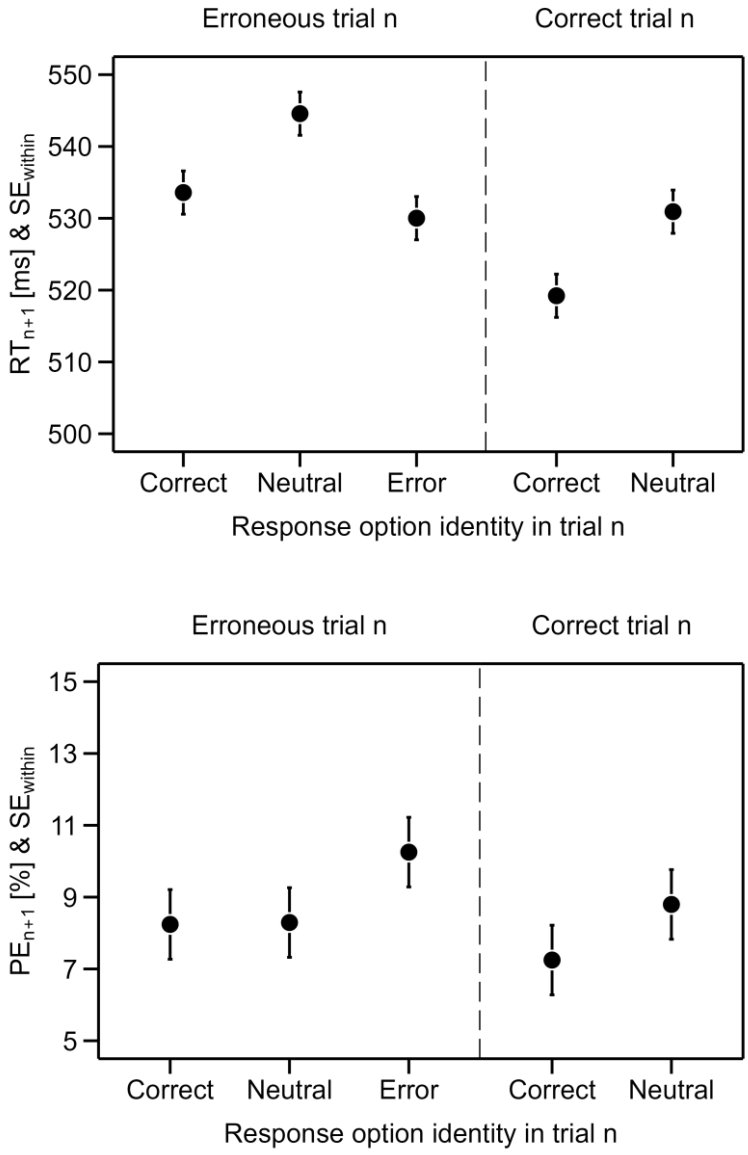
In sum, after an error in trial n, measures of sub-threshold muscle activity during the pre-stimulus interval of trial n+1 provided no evidence for (i) response-specific activation or inhibition of either the erroneous response or the correct response or (ii) response-general inhibition or conservatism.

After error commission in trial n+1: Responding

In H6 and H7, we investigated changes in the initiation of overt responses in trial n+1. H6 targeted response-specific activation or inhibition. H7 targeted response-general inhibition or conservatism (see Fig. 5). We considered these preregistered analyses to be exploratory because we expected, and therefore accepted, (i) fewer observations per participant and (ii) additional participant exclusions (see *Data treatment*).

Figure 5

Results for H6 and H7 in Experiment 1



Note. Response time in trial n+1 (RT_{n+1}, top) and percentage errors in trial n+1 (PE_{n+1}, bottom). The left side of the figure shows the correct, neutral, and erroneous response from an erroneous trial n. The right side of the figure shows the correct and neutral response from a correct trial n. The error bars depict the within-subject standard error for the full factorial design (SE_{within}; Loftus & Mason, 1994).

H6 predicts that RT_{n+1} and PE_{n+1} will differ between the three response options after error commission in trial n, indicating response-specific changes. Inhibition of the erroneous response should increase RT_{n+1} for the erroneous response from trial n compared to the correct or neutral response from trial n. Further, errors should be committed less frequently (i.e., PE_{n+1} should decrease) by selecting the erroneous response than the other two responses from trial n. Note, here, that PE_{n+1} therefore denotes the frequency of committing an error by selecting a response of a certain identity from trial n. In contrast,

activation of the correct or the erroneous response after error commission should decrease RT_{n+1} for the correct or the erroneous response from trial n compared to the neutral response from trial n. Further, errors should be committed more frequently (i.e., PE_{n+1} should increase) by selecting the correct or the erroneous response from trial n than by selecting the neutral response from trial n.

To test these predictions, we analyzed RT_{n+1} and PE_{n+1} after erroneous trials in separate ANOVAs with a single within-subjects factor: response option identity in trial n (correct vs. neutral vs. erroneous). A significant main effect emerged for RT_{n+1}, $F(2, 88) = 6.56$, $p = .002$, $\eta_p^2 = .13$. To scrutinize this main effect, we compared the three responses from trial n with two-tailed paired-samples *t*-tests. Consistent with response-specific activation of the correct and the erroneous response, RT_{n+1} for the neutral response was slower than for both the correct response, $t(44) = 2.56$, $p = .014$, $d_z = 0.38$, and the erroneous response, $t(44) = 4.06$, $p < .001$, $d_z = 0.61$. In contrast, there was no difference in RT_{n+1} between the correct and erroneous response, $|t| < 1$. For PE_{n+1}, however, the main effect of response option identity in trial n was not significant, $F(2, 88) = 1.38$, $p = .258$, $\eta_p^2 = .03$.

H7 predicts that RT_{n+1} will be higher and that PE_{n+1} will be lower for the neutral response after an erroneous trial n as compared to after a correct trial n, indicating response-general inhibition or conservatism after committing an error. We confirmed this hypothesis in a one-tailed, paired-samples *t*-test for RT_{n+1}, $t(44) = 5.48$, $p < .001$, $d_z = 0.82$, but not for PE_{n+1}, $|t| < 1$.

In sum, analyses of RT_{n+1} provided evidence for both (i) response-specific activation of the erroneous and correct responses and (ii) response-general inhibition or conservatism. Analyses of PE_{n+1}, however, yielded only null effects. The latter data pattern is not consistent with our hypotheses but also does not contradict them. We simply cannot draw strong conclusions from a null effect.

We also preregistered to explore the same comparisons as in H6 and H7 for RD_{n+1} and $MF_{RD_{n+1}}$ with the goal of investigating changes in the execution of overt responses in trial n+1. We had no specific hypotheses for these comparisons. The analyses of these dependent variables (which are presented in *Appendix C*) did not reveal any reliable impact of response-specific or response-general processes.

Discussion

In Experiment 1, we distinguished between response-specific and response-general processing changes in the context of commission errors. More specifically, we investigated these processing changes during (i) the commission of an error, (ii) the pre-stimulus interval of trial n+1, and (iii) the subsequent response. Replicating prior findings, our data support the view that response-specific error

cancellation and correction efforts occur during the commission of hastily initiated errors (e.g., Bode & Stahl, 2014; Foerster, Steinhauser, et al., 2022; Hochman et al., 2017; Rabbitt, 1966). Our data further indicate that, following an error in trial n, there is no evidence for response-specific activation or inhibition of the erroneous or correct response in sub-threshold muscle activity measured during the pre-stimulus interval of trial n+1. Our data also show that, following an error in trial n, there is evidence for response-specific activation of the correct response and the erroneous response during response selection in trial n+1. Finally, consistent with response-general processes, we observed slowing of the neutral response from trial n in trial n+1 after participants made an error (vs. a correct response) in trial n. These findings indicate post-error changes in overt responses that are both response-specific (e.g., increased correct response activation) and response-general (i.e., generalized inhibition or more conservative responding) in nature. This outcome is important because the response-specific processing changes that we have identified may contribute to post-error changes in performance that are typically attributed solely to response-general processing changes.

There are several possible explanations for why response activation after error commission is absent in the pre-stimulus interval but present in subsequent responding. First, sub-threshold force might not be sensitive enough to capture response activation that carries over from trial n to trial n+1 because (i) there is too much noise for too little signal or (ii) cognitive activation does not directly leak into sub-threshold motor activation during the time between trials. In other words, response activation from the erroneous trial may have been carried over into the next trial, although it did not show up in the $MF_{PreS_{n+1}}$ measurement. Multiple previous studies successfully detected modulations of response activation in sub-threshold response force shortly before the onset of task-relevant stimuli with similar or even smaller sample sizes (Weissman, 2019; Weissman & Schmidt, 2024; Weissman et al., 2024). However, these studies revealed response activations based on expectations about upcoming responses. Speculatively, only strongly anticipated responses might leak into traceable, sub-threshold motor activity, and errors might not come with such biased expectations about the upcoming response.

Second, both responses might be re-activated through the retrieval of bindings (e.g., Frings et al., 2024; Frings et al., 2020). Episodic binding and retrieval creates short-cuts in action control for errors between features of stimuli, correct and erroneous responses, and their effects (Foerster et al., 2023; Foerster, Moeller, et al., 2022; Foerster et al., 2021). However, in our design, stimuli always changed in consecutive trials. As such, a response could not be retrieved from the stimulus it was previously bound to because a stimulus never repeated from trial n to trial n+1. A binding and retrieval mechanism may

still be at work, though, considering recent evidence that responses are bound to and retrieved from the task in which they were performed (Benini et al., 2024). As the task was always the same in the current experiment, the repetition of the task in each trial could have triggered the retrieval of the previously bound correct and erroneous responses after an error even without a stimulus or response repetition. Retrieval of a response would facilitate its performance, explaining faster RTs after errors for correct and erroneous responses compared to neutral responses. However, if a task repetition is registered upon letter onset, response retrieval would not affect behavior during the pre-stimulus interval, explaining the null results we observed in pre-stimulus force. Whether the activation of correct and erroneous responses is due to retrieval from bindings with the task can be tested in a design that manipulates task sequence from trial n to trial n+1: Correct and erroneous response activation in trial n+1 should emerge more strongly if the task repeats than if the task changes.

Finally, it is possible that there is specific or general response inhibition followed by a rebound of activation for the correct and erroneous responses. This rebound could occur when the representations of the two responses are re-activated during the evaluation of the action, which likely involves assessing what went wrong and what would have been right (Jentzsch & Dudschig, 2009; Mohr et al., 2018; Wirth et al., 2018). Yet, sub-threshold motor activation in the pre-stimulus interval neither supports an initial inhibition nor a rebound of activation. Therefore, to explain our findings, this explanation would need to posit that the force measurement during this interval is not sensitive enough to detect these processes.

Unplanned, exploratory analyses

We have argued that faster RTs in trial n+1 for the neutral response from a correct versus erroneous trial n indicate response-general processes (e.g., generalized inhibition). There is, however, another potential explanation for this effect that involves response-specific processes. Heightened activation of both the correct response and the erroneous response following an erroneous trial n may lead to heightened response conflict when selecting the neutral response from trial n in trial n+1. Such conflict could slow RTs, consistent with prior findings indicating that conflict has a detrimental effect on performance across different conflict inducing tasks (e.g., Kornblum et al., 1990; Simon & Rudell, 1967; Stroop, 1935). In contrast, if only the correct response remains highly active after a correct trial n, then competition with the neutral response would stem from only one alternative response after correct trials rather than from two alternative responses after erroneous trials. Therefore, differences in performance involving the neutral response between post-correct and post-error trials might not solely reflect

response-general processes (e.g., a shifted response threshold or generalized inhibition). They might additionally (or instead) reflect response-specific processes that produce varying degrees of response conflict during response selection after an error, which lead to mutual inhibition of competing response alternatives.

We explored this competition hypothesis further in unplanned exploratory analyses before we tested it a priori in Experiment 2. Our rationale was as follows. The presence of response-specific activation after an erroneous trial n strongly suggests that there is also response-specific activation of the correct response after a correct trial n. Therefore, following a correct trial n, performance in trial n+1 should be better for the correct response than for the neutral response. This advantage could stem from heightened activation of the correct response or competition-induced slowing of the neutral response. Further, if only the correct response is specifically activated after a correct trial n, but both the correct and the erroneous response are activated after an erroneous trial n, then the correct response following a correct trial n would not encounter conflict. In contrast, the correct response after an erroneous trial n would conflict with the concurrently activated erroneous response. Therefore, performance in trial n+1 should be better for the correct response from a correct trial n than for both the correct response and the erroneous response from an erroneous trial n.

To comprehensively assess the ex-post facto hypothesis of competing activation of the correct and erroneous response in trial n+1, we analyzed differences in RT_{n+1} and PE_{n+1} between all combinations of response option identity and response accuracy in trial n (see Fig. 5). Specifically, we conducted two separate ANOVAs with a single within-subjects factor that allowed for a comparison of all five combinations of response option identity and response accuracy (i.e., correct response from an erroneous trial, neutral response from an erroneous trial, erroneous response from an erroneous trial, correct response from a correct trial, and neutral response from a correct trial). For completeness, we conducted an analogous analysis for $MF_{PreS_{n+1}}$, which we report in *Appendix D*. For the condition involving a correct response from a correct trial n, which was previously not analyzed and therefore not considered in the original outlier detection, we excluded additional outliers in RT_{n+1}, RD_{n+1}, and $MF_{RD_{n+1}}$ (1.9%). We included RD_{n+1} and $MF_{RD_{n+1}}$ because we also conducted complementary exploratory analyses for these dependent measures (see *Appendix C*).

In RT_{n+1}, we found a significant difference between conditions, $F(4, 176) = 12.51$, $p < .001$, $\eta_p^2 = .22$ (Greenhouse-Geisser corrected; $\varepsilon = .72$). As we reported in our investigation of H6, RT_{n+1} was faster for both the correct and erroneous response than for the neutral response from an erroneous trial n, but

no significant difference emerged for the comparison of the correct and erroneous response. The analysis for H7 already revealed that RT_{n+1} for the neutral response was higher after an erroneous trial n than after a correct trial n. We analyzed the remaining six comparisons in two-tailed paired-samples *t*-tests. Consistent with the response competition hypothesis, RT_{n+1} for the correct response from a correct trial n was always faster than for the other four conditions (i.e., each of the three responses from an erroneous trial n and for the neutral response from a correct trial n), all *t*s ≥ 3.21, $p \leq .003$, $d_z \geq 0.48$. Comparing the neutral response from a correct trial n to the erroneous and correct responses from an erroneous trial n yielded no significant effects, both |*t*|s < 1. Finally, there were no significant differences between the five conditions in PE_{n+1}, $F(4, 176) = 1.68$, $p = .173$, $\eta_p^2 = .04$ (Greenhouse-Geisser corrected; ε = .76).

In sum, consistent with activation of the correct response and the response competition hypothesis, the exploratory analyses revealed faster RT_{n+1} for the correct response than for the neutral response after a correct trial n. Critically, we also observed faster RTs for the correct response from a correct trial n than for the correct and erroneous responses from an erroneous trial n, consistent with response competition between the latter two responses after an erroneous trial n.

Experiment 2

Introduction

An important limitation of Experiment 1 is that the crucial findings indicating response-specific activation and competition in RT_{n+1} came from exploratory analyses. Thus, the aim of Experiment 2 is to replicate these findings. For this endeavor, the measurement of RT is critical, whereas a fine-grained assessment of response force is no longer necessary. Therefore, to streamline data collection, we assessed performance via a standard keyboard instead of with force-sensitive keys using an otherwise identical paradigm. This reduced setup allows us to replicate poor-response planning (H1) and cancellation (H2) during error commission, response-specific activation (H3), response-specific competition, as well as response-general inhibition or conservatism for response selection following error commission (H4). The hypotheses will again be specified in the *Results* section (see Table 2).

Method

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. We preregistered this study before data collection (osf.io/f49z7). We collected the data in 2025. We made the program code, data, and analysis code publicly available (osf.io/f49z7). We processed and analyzed the data using the same software as in Experiment 1.

Participants

The critical effect sizes from Experiment 1 were $d_z \geq 0.38$ in RT, RD, and RT_{n+1} (see H1 to H4 in Table 2). At an alpha of .05, a sample of 60 participants provides a power of ≥ .90 to detect $d_z \geq 0.38$ in a one-tailed, paired-samples *t*-test.

Sixty participants who provided written informed consent and received monetary compensation or partial course credit were included in the final sample (mean age = 25, *SD* = 5.7; 42 female, 18 male, none non-binary; 50 right-handed, seven left-handed, three ambidextrous). Data from 19 additional participants were collected but excluded from all analyses and replaced with other participants for the following reasons. First, although we did not include the following three scenarios in the preregistration, we had to exclude and replace several datasets because their sessions deviated in one of three ways from the intended procedure: Seven participants were invited to the second session but did not schedule a second session, one participant was accidentally assigned different stimulus-response mappings in their two sessions, and one participant used only two of the three instructed keys in both sessions. Second, as preregistered, eight participants were not invited to the second session because they responded correctly in less than 50% of the trials during the first session. Similarly, an additional participant responded correctly in less than 50% of the trials during the second session and was also excluded. Third, as preregistered, one participant was excluded because their data provided fewer than 5 observations in a design cell that we needed to test the hypotheses (see *Data treatment*).

Stimuli, apparatus and procedure

Experiment 2 was the same as Experiment 1, but participants responded through a full-size computer keyboard. We simplified the response setup because keypress responses were sufficient to test the hypotheses of Experiment 2. Further, relying on keyboards rather than a single custom force

key apparatus allowed for more efficient data collection from up to five participants at the same time. Participants put their index, middle, and ring finger on the adjacent leftwards, downwards, and rightwards pointing arrow keys. To make the experiments as similar as possible, we defined response on- and offsets as in Experiment 1 and sampled the status of each key at 250 Hz. Because the status of the key is dichotomous (i.e., pressed down vs. not pressed down), there was no baseline measurement or individual response threshold. Akin to Experiment 1, response onset was defined as the first instance of five consecutive measurements during which a key was pressed down. Response offset was defined as the first instance thereafter where the key was not pressed down and kept this status for another four consecutive measurements.

Data treatment

In contrast to Experiment 1, we did not collect or analyze force data. For this reason, we did not have to resample the data as in Experiment 1 (we sampled effectively every 4.2 ms, *SD* = 0.03 ms). As in Experiment 1, however, we excluded several types of trials. First, we excluded the trials in the practice block of each session. Second, we excluded the first and last trial of each experimental block. Third, we excluded trials with an error in the preceding trial n-1 (22.2%).

We selected trial sequences with either a correct response or a commission error in trial n and in trial n+1. Therefore, we excluded trial sequences with early responses during fixation (< 0.1%), omission errors (5.8%), and additional responses during the blank screen (2.6%) in trial n. We further excluded trial sequences with early responses during fixation (< 0.1%), omission errors (5.6%), and additional responses during the blank screen (2.5%) in trial n+1. For this selection of trial sequences, we computed PE_{n+1}.

For the analyses of all other dependent measures (i.e., all measures except for PE_{n+1}), we made the following additional exclusions. First, we excluded trial sequences with commission errors in trial n+1 (11.9%). Second, we excluded outliers separately for each participant, dependent measure (RT, RD, and RT_{n+1}), and corresponding design cell of the planned analyses (4.3%).

After these exclusions, one participant had less than 5 observations in a design cell for the analyses related to H1-H5 and was excluded and replaced (see *Participants*).

Results

In the following, we again differentiate the analyses with respect to their focus on processes (i) *during* error commission (H1 and H2) or (ii) *after* error commission during the response of trial n+1 (H3 and H4). We report the results of planned, secondary analyses after the results for the hypotheses.

Table 2 provides a summary of the assumed processes for the different time points and hypotheses, hypothesized effects and empirical findings. The two central factors in all our analyses are again (i) *response accuracy in trial n* and (ii) *response option identity in trial n*.

Table 2

Timing, processes, hypotheses, and outcome of Experiment 2

Error process	H	DV	Hypothesized effect	Outcome
			<u>During error comission in trial n</u>	
Poor pre-response planning	1	RT	erroneous trial n < correct trial n	✔
Response-specific cancellation	2	RD	erroneous trial n < correct trial n	✔
			<u>After error commission in trial n+1: Responding</u>	
Response-specific activation	3a	RT_{n+1}	correct/erroneous response < neutral response from erroneous trial n	✔
	3b	RT_{n+1}	correct response < neutral response from correct trial n	✔
Response-specifc competition and response-general inhibition or conservatism	4a	RT_{n+1}	correct response from correct trial n < correct/erroneous response from erroneous trial n	✔
	4b	RT_{n+1}	neutral response from erroneous trial n > from correct trial n	✔

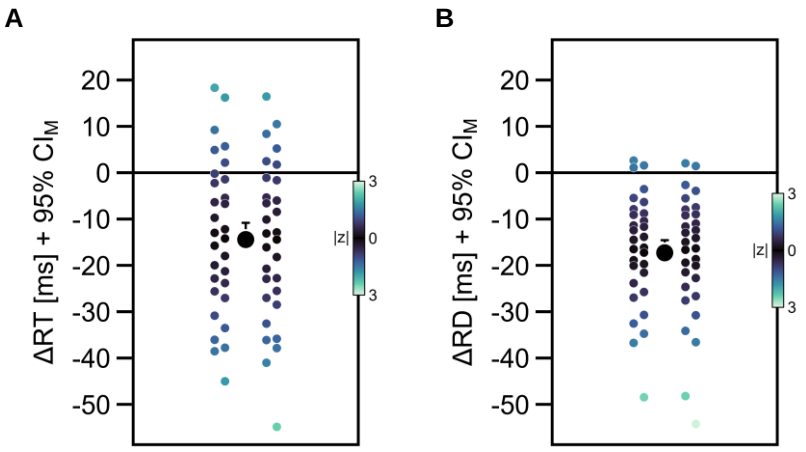
Note. Overview of hypotheses (H) 1-4 to test error processes assumed to operate during error commission or after error commisson during responding. The table further specifies the hypothesized effects for each dependent variable (DV). A check mark indicates that the data produced an outcome that is consistent with the hypothesized effect.

During error commission in trial n

In H1 and H2, we investigated processes during error commission. Specifically, H1 targeted poor pre-planning of errors, whereas H2 targeted the response-specific process of error cancellation (see Fig. 6).

Figure 6

Results for H1 and H2 in Experiment 2



Note. Differences (Δ ≔ erroneous trial n – correct trial n) in **(A)** response time (RT; H1) and **(B)** response durations (RD; H2). Large, black dots indicate mean differences. Error bars visualize a one-tailed *t*-test. They correspond to the upper bound of the 95% confidence interval of the related difference between means (CI_M). Small dots indicate individual participant values, and color visualizes the standardized deviation from the mean difference (|z|).

H1 predicts that erroneous responses will be initiated faster than correct responses due to less pre-response planning. Consistent with this hypothesis, a one-tailed, paired-samples *t*-test revealed significantly faster RT in erroneous trials than in correct trials (497 vs. 511 ms; see Fig. 6A), $t(59) = -6.68$, $p < .001$, $d_z = -0.86$.

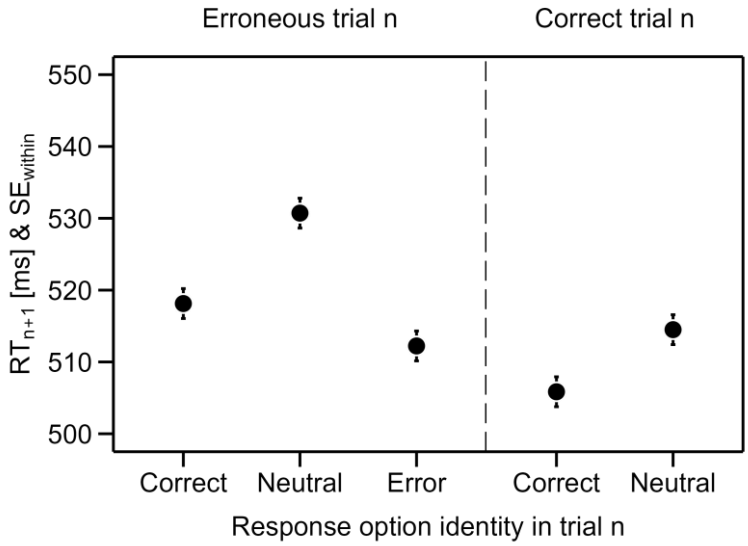
H2 predicts that the execution of an erroneous response will be shorter than the execution of a correct response due to weaker initiation and active cancellation of erroneous responses. Consistent with this hypothesis, a one-tailed, paired-samples *t*-test revealed significantly shorter RD in erroneous trials than in correct trials (117 vs. 135 ms; see Fig. 6B), $t(59) = -10.60$, $p < .001$, $d_z = -1.37$.

After error commission in trial n+1: Responding – Primary analyses

In H3 and H4, we investigated changes in the initiation of overt responses in trial n+1. H3 targeted response-specific activation and H4 targeted response-specific competition and response-general inhibition or conservatism (see Fig. 7). As an initial step, we analyzed RT_{n+1} in an ANOVA with a single within-subjects factor to compare all five combinations of response accuracy in trial n and response option identity in trial n (i.e., correct response from an erroneous trial, neutral response from an erroneous trial, erroneous response from an erroneous trial, correct response from a correct trial, neutral response from a correct trial). As the main effect was significant, $F(4, 236) = 26.12$, $p < .001$, $\eta_p^2 = .31$ (Greenhouse-Geisser corrected; $\varepsilon = .75$), we tested H3 and H4 in six one-tailed, paired-samples *t*-tests.

Figure 7

Results for H3 and H4 in Experiment 2



Note. Response time in trial n+1 (RT_{n+1}). The left side of the figure shows the correct, neutral, and erroneous response from an erroneous trial n. The right side of the figure shows the correct and neutral response from a correct trial n. The error bars depict the within-subject standard error for the full factorial design (SE_{within}; Loftus & Mason, 1994).

H3 concerns response-specific activation of the correct response or the erroneous response from trial n in trial n+1. H3a predicts that, after an erroneous trial n, the initiation of both the correct response and the erroneous response will be faster in trial n+1 than the initiation of the neutral response due to stronger activation of the former two responses. Consistent with this hypothesis, RT_{n+1} was significantly faster for both the correct response, $t(59) = -4.21$, $p < .001$, $d_z = -0.54$, and the erroneous response, $t(59) = -6.09$, $p < .001$, $d_z = -0.79$, than for the neutral response.

H3b predicts that, after a correct trial n, the initiation of the correct response from trial n in trial n+1 will be faster than the initiation of the neutral response due to stronger activation of the executed response. Consistent with this hypothesis, RT_{n+1} was significantly faster for the correct response than for the neutral response, $t(59) = -5.15$, $p < .001$, $d_z = -0.67$.

H4 concerns the influence of both response-specific and response-general processes on performance in trial n+1. First, it predicts that response-specific activation of the correct response and the erroneous response after an erroneous trial n will increase response competition compared to the activation of only the correct response after a correct trial n. Second, it predicts that response-general processes will slow the initiation of all responses after an erroneous trial n.

H4a predicts that, in trial n+1, the initiation of the correct response from a correct trial will be faster than the initiation of both the erroneous response and the correct response from an erroneous trial n, indicating response-specific competition between the erroneous and correct response and/or response-general slowing after errors. Consistent with this hypothesis, RT_{n+1} was significantly faster for the correct response from a correct trial n than for both the correct response, $t(59) = -5.21$, $p < .001$, $d_z = -0.67$, and the erroneous response, $t(59) = -2.60$, $p = .006$, $d_z = -0.34$, from an erroneous trial n.

H4b predicts that, in trial n+1, the initiation of the neutral response from trial n will be faster after a correct trial n than after an erroneous trial n. Such an effect could indicate stronger competing response activation from two responses after an error instead of one response after a correct trial and/or response-general slowing after errors. Consistent with this hypothesis, RT_{n+1} was significantly faster for the neutral response from a correct trial n than from an erroneous trial n, $t(59) = -9.01$, $p < .001$, $d_z = -1.16$.

In sum, our findings support both H3 and H4. While the results for H3 clearly suggest the operation of response-specific processes, the results for H4 are in line with (i) response-general processes and/or (ii) response competition arising from distinct response-specific activations after errors.

There are some remaining comparisons between conditions for RT_{n+1} that we did not include in H3 and H4 because they do not provide additional insight. For completeness, we report the results of these comparisons in *Appendix E*.

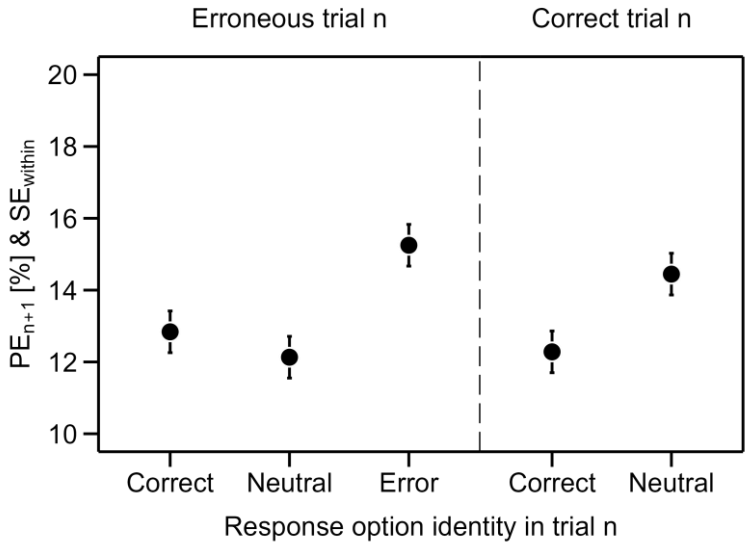
After error commission in trial n+1: Responding – Secondary analyses

Analogous predictions as for RT_{n+1} can be made for PE_{n+1}. However, because PE_{n+1} is generally less sensitive than RT_{n+1} in error research, we did not expect to find reliable effects in this measure (see Exp. 1). Therefore, we analyzed PE_{n+1} only as a secondary measure, with response option identity denoting which response from trial n was executed as an error in trial n+1 (see Fig. 8). The main effect in the ANOVA was significant, $F(4, 236) = 6.82$, $p < .001$, $\eta_p^2 = .10$ (Greenhouse-Geisser corrected; $\varepsilon = .84$). Accordingly, we conducted pairwise comparisons via two-tailed paired-samples *t*-tests[3]. As in Experiment 1, PE_{n+1} denotes the frequency of committing an error by selecting a response of a certain identity from trial n.

[3] We conducted two-tailed rather than one-tailed *t*-tests because we did not formulate directional hypotheses for this measure.

Figure 8

Results for PE_{n+1} in Experiment 2



Note. Percentage errors in trial n+1 (PE_{n+1}, bottom). The left side of the figure shows the correct, neutral, and erroneous response from an erroneous trial n. The right side of the figure shows the correct and neutral response from a correct trial n. The error bars depict the within-subject standard error for the full factorial design (SE_{within}; Loftus & Mason, 1994).

First, PE_{n+1} did not differ significantly between the correct response and the neutral response from an erroneous trial n, $|t| < 1$. However, participants committed more errors in trial n+1 by selecting the erroneous (vs. neutral) response from an erroneous trial n, $t(59) = 3.88$, $p < .001$, $d_z = 0.50$. The latter result suggests that the erroneous response was more active than the neutral response. This finding partly corroborates (for the erroneous response) and does not contradict (for the correct response) response-specific activation after an erroneous trial n, as suggested by the findings of H3a.

Second, after a correct trial n, participants committed fewer errors in trial n+1 by selecting the correct response from trial n than by selecting one of the neutral responses from trial n, $t(59) = -4.68$, $p < .001$, $d_z = -0.60$. This result suggests that the correct response was less active than the neutral response. This finding appears to contradict response-specific activation of the executed correct response following a correct trial n, as suggested by the findings of H3b.

Third, participants committed errors in trial n+1 less frequently by selecting the correct response from a correct trial n than by selecting the erroneous response from an erroneous trial n, $t(59) = -3.93$, $p < .001$, $d_z = -0.51$. However, PE_{n+1} for the correct response from a correct trial n did not differ significantly from PE_{n+1} for the correct response from an erroneous trial n, $t(59) = -0.66$, $p = .513$, $d_z = -0.08$. The findings therefore partly (for the erroneous response) and partly do not (for the correct

response) contradict response-specific competition and response-general processes after an erroneous trial n, as suggested by the findings of H4a.

Fourth, participants committed an error in trial n+1 more frequently by selecting the neutral response from a correct trial n than from an erroneous trial n, $t(59) = 3.66$, $p = .001$, $d_z = 0.47$. This finding corroborates response-specific competition or response-general processes after an erroneous trial n, as suggested by the findings of H4b.

As for RT_{n+1}, there are some remaining comparisons that do not provide additional insight into our main hypotheses. For completeness, however, we report the results of these tests in *Appendix E*.

Discussion

Experiment 2 replicated the key findings from Experiment 1. First, it replicated our findings indicating that error cancellation is a response-specific process that occurs during the commission of a hastily initiated error (e.g., Bode & Stahl, 2014; Foerster, Steinhauser, et al., 2022; Hochman et al., 2017; Rabbitt, 1966). Second, it replicated the critical exploratory findings from Experiment 1 related to response-specific and response-general processing changes in response selection during trial n+1. Indeed, the analysis of RT_{n+1} confirmed all our hypotheses and thus fully replicated the pattern from Experiment 1. First, the evidence points to response-specific activation. Namely, the correct response and the erroneous response appear to be more active than the neutral response after an erroneous trial n, as indicated by faster RT_{n+1} for the former responses than for the latter one. Further, the correct response from a correct trial n appears to be more active than the neutral response from a correct trial n, as indicated by faster RT_{n+1} for the correct response than for the neutral response. Second, the evidence points to response-specific competition and/or response-general inhibition or conservatism. The correct and erroneous responses from an erroneous trial n seem to compete with each other or to be slowed down together with all responses, as indicated by longer RT_{n+1} relative to the correct response from a correct trial n. Further, the neutral response seems to compete with the two other responses, or it also is subject to a general slowdown, as indicated by longer RT_{n+1} for the neutral response from an erroneous trial n than from a correct trial n.

The secondary analyses of PE_{n+1}, however, yielded mixed results with respect to H3 on response-specific activation. After an erroneous trial n, the results align with the findings from RT_{n+1} related to response-specific activation (H3a): Specifically, increased activity of the erroneous response is suggested by higher PE_{n+1} compared to the neutral response, and the correct response and the neutral

response did not differ in PE_{n+1}. After a correct trial n, however, the results diverge from the findings from RT_{n+1} related to response-specific activation (H3b): Specifically, the correct response was initiated faster, suggesting higher activation, but it was also selected less frequently when committing an error, suggesting lower activation.

With respect to H4 concerning response-specific competition and response-general inhibition or conservatism, the results from PE_{n+1} again align only partly with the results from RT_{n+1}. In line with all proposed processes, the neutral response was less frequently selected as error in trial n+1 if trial n was erroneous rather than correct. Thus, for H4b, the findings for PE_{n+1} corroborate the findings for RT_{n+1}. However, participants committed more errors by selecting the erroneous response from an erroneous trial n than by selecting the correct response from a correct trial n. Here, we would have expected the opposite effect because (i) competition between the erroneous response and the correct response from an erroneous trial n should reduce the erroneous selection of both responses in trial n+1 and/or (ii) a more conservative response criterion or generalized inhibition should decrease error commission generally after an erroneous trial n.

The PE_{n+1} data thus support the RT_{n+1} results for H3a and H4b (which correspond to H6 and H7 in Experiment 1) but contradict the RT_{n+1} results for H3b and H4a (which were derived from the unplanned, exploratory analyses of Experiment 1). Notably, these findings are not readily explained by conflict arising from response-specific activations, nor by response-general processes. Especially the unexpected pattern for the correct response from a correct trial n suggests that other response-specific processes beyond response-specific activation also contribute to post-error and post-correct behavioral changes. We currently lack a theoretical explanation for this pattern of results.

But why did Experiment 2 reveal effects in PE_{n+1}, whereas Experiment 1 yielded null results? First, a visual comparison of Figures 5 and 8 suggests a similar descriptive pattern of results across experiments. Second, this visual comparison also suggests an overall higher error rate in Experiment 2. Albeit speculative, Experiment 1 might therefore have lacked the sensitivity to reliably measure changes in accuracy due to a floor effect.

In sum, Experiment 2 strengthened the exploratory findings from Experiment 1 related to response-specific activation of the correct and erroneous response during responding in trial n+1 after error commission in trial n. Further, the primary RT analyses and part of the secondary accuracy analyses also support the assumption that response-specific competition and response-general processes (i.e., generalized inhibition or more conservative responding) contribute to post-error

performance. Finally, the secondary accuracy analyses partly contradict the latter assumption and post-correct response activation. However, we argue in the *General Discussion* that previous results on post-error accuracy have also been mixed.

General Discussion

Many prominent theories implicitly assume that post-error performance reflects only response-general processes (e.g., equivalent slowing of all possible responses). The present findings challenge this implicit assumption by showing that response-specific processes also influence post-error performance. Indeed, data from two experiments suggest that error commission in trial n leads to increased activity of the correct and the erroneous response during response selection in trial n+1. Our data also suggest that, after an error, responding in trial n+1 is slowed by (i) response competition arising from the selective activation of the correct and erroneous response and/or (ii) response-general processes like generalized inhibition. These findings have important implications for theories of post-error processing including those that involve conflict monitoring and generalized inhibition. We discuss the implications of our findings below.

A novel perspective on post-error processing

Computational models derived from conflict monitoring theory offer an integrative account of error processing by regarding errors as an instance of behavioral conflict. Our observation that both the correct response and the erroneous responses are more active than the neutral response in trials after an error, however, contradicts current model predictions. Indeed, model simulations demonstrate that response conflict between the correct and erroneous responses is high after an error but is resolved within a few hundred milliseconds through mutual inhibition of active response representations (Yeung et al., 2004). Thus, subsequent performance is affected by conflict only in a response-general manner: After actions with high conflict, cognitive control is increased, thereby inducing a more conservative speed-accuracy tradeoff for all actions (e.g., Botvinick et al., 2001). In contrast, our findings suggest that response-specific activation (and any associated mutual inhibition) can be evident over an extended period, exerting a direct impact on upcoming actions (see also Adkins et al., 2024, for recent evidence on prolonged activation of erroneous responses). A suitable revision of the conflict monitoring model might involve a weaker or nonlinear inhibitory connection between response representations, thereby allowing a certain amount of response activation to persist over time.

The cognitive processes that unfold after an error have been described as a sequence of maladaptive processes followed by adaptive processes (e.g., Gjorgieva & Egner, 2022; Wessel, 2018). Therefore, post-error changes in performance should reflect maladaptive processes (e.g., generalized inhibition and orienting of attention away from the task) that occur directly after an error, but adaptive processes (e.g., conservative criterion adjustment) when more time has passed (e.g., Jentzsch & Dudschig, 2009). The current study suggests that such a chronological order does not incorporate (i) early processes that should be regarded as adaptive or (ii) late processes that are to some extent maladaptive. Regarding early adaptive processes, the current results replicate recent findings related to error cancellation, a clearly adaptive process that operates during the commission of an error (Foerster, Steinhauser, et al., 2022). Regarding late maladaptive processes, our results indicate increased accessibility of both the correct response and the erroneous response during response selection following an error. This outcome could be construed as adaptive if it reflects – or is a byproduct of – binding the correct and erroneous responses to tasks, stimuli, or effects as described earlier. Even in this situation, however, our findings suggest that increased accessibility of these responses would likely be a source of interference that hampers the selection of other (neutral) actions. Therefore, the present findings challenge the popular view that post-error processes reflect a gradual transition from solely maladaptive processes to solely adaptive processes over time.

We examined conflicting response activation after an error only for a relatively long response-stimulus interval, which is when adaptive post-error processes should be active (e.g., Gjorgieva & Egner, 2022). Therefore, it is an open question whether we would observe the same pattern in a timeframe for which maladaptive processes are still predominant (i.e., when the response-stimulus interval is short or nonexistent; e.g., Jentzsch & Dudschig, 2009; M. Steinhauser et al., 2017; R. Steinhauser & Steinhauser, 2021). Observing such activation of the correct and erroneous response across a timespan that was previously assumed to distinguish between maladaptive and adaptive processes would further strengthen our view that a strict chronological order of these processes is not warranted. We suspect that both responses would indeed modulate performance after an error even for nonexistent response-stimulus intervals (i.e., 0 ms), because of the recent execution of the erroneous response and fast activation of the correct response, as demonstrated in our study and previous studies (e.g., Rabbitt, 1966, 2002). Further, substantially longer response-stimulus intervals could reveal the durability of these activations. A systematic manipulation of the temporal interval between successive trials would therefore

allow assessing the time course of the impact of correct and erroneous response activation on response selection (Jentzsch & Dudschig, 2009; M. Steinhauser et al., 2017).

Limitations

A strict chronological order of early maladaptive and late adaptive processes offers a clear prediction about the accuracy of responses after errors (Wessel, 2018). Maladaptive processes such as increased monitoring of an error should make subsequent responses more error-prone when the response-stimulus interval is sufficiently short. Adaptive processes should instead improve accuracy after a sufficiently long response-stimulus interval. Studies indeed observed consistently lower accuracy after errors than after correct responses for short intervals in line with maladaptive processes (Jentzsch & Dudschig, 2009; M. Steinhauser et al., 2017). However, the accuracy data for long intervals did not support adaptive processes because post-error responses either did not differ from post-correct responses (Jentzsch & Dudschig, 2009; M. Steinhauser et al., 2017) or they were still more error-prone (M. Steinhauser et al., 2017). We likewise did not find differences in accuracy between response option identities after errors or between neutral responses after erroneous compared to correct trials in Experiment 1. In Experiment 2, the accuracy data partly aligns with adaptive error processing (i.e., fewer error commissions in trial n+1 with the neutral response after erroneous than correct trials n; see Fig. 8) but also partly contradicts it (i.e., committing the same erroneous response in trial n and trial n+1 was amongst the most frequent errors). Response-specific and response-general processes, at least the ones proposed in the current study, also cannot explain the whole pattern of results. Although this research endeavor carved out novel, error-related processes, it also reveals that current theories do not explain the entirety of post-error changes.

As we described earlier, explaining our findings related to response-specific processes in terms of residual, rebound, or retrieved activation of the erroneous and correct response provides a parsimonious model for both the current results and previous findings. We acknowledge, however, that the present observations could stem from the superposition of multiple independent cognitive and/or motoric processes. For example, the data pattern in the response times of trial n+1 could be explained by a combination of efficient re-activation of previous motor patterns and response-specific inhibition. With regard to re-activation of previous motor patterns, the faster initiation of the correct response after a correct trial n might originate from facilitated re-contraction of previously engaged muscles. This benefit should, however, be similar for the erroneous response after erroneous trials. It should also be

evident, albeit to a smaller extent, for correct responses from an erroneous trial n, as our measures of PF in Experiment 1 suggest sub-threshold motor activity for correct responses in these trials. Accordingly, response-specific inhibition would need to target both (i) the erroneous response (counteracting benefits of motor repetitions) and (ii) the neutral response to fully account for the observed RT_{n+1} pattern after errors.

Conclusion

Current views posit that post-error performance reflects response-general processes that operate on all cognitive representations of responses of a task (e.g., a shift in the decision criterion), rather than processes that operate on specific responses. Contrary to these views, our findings indicate selective activation during response selection in trial n+1 of both the correct response and the erroneous response (vs. the neutral response) from an erroneous trial n. Our findings further suggest that the simultaneous activation of the correct and erroneous responses after an error engenders a cognitive conflict that influences response selection in trial n+1. By revealing that response-specific processes influence post-error performance in these ways, our findings challenge current views of error processing that posit only response-general processes. In particular, our findings show that considering specific response identities in designs with more than two response options allows for a more precise description of post-error behavior than does averaging across all responses, as in traditional measures of post-error performance. To separate response-specific influences from other error-related processes in future studies of error processing, researchers should consider the identity of the current response in relation to the previous response.

Author contributions

All authors developed the concept and methodology. AF, RP, and WK acquired funding and provided the resources. AF supervised the research, administered the project, and programmed the experiment. AF and MS organized the investigation, curated, validated, and analyzed the data, visualized the method and results, and wrote the first draft of the manuscript. All authors reviewed and edited the manuscript.

Appendices

Appendix A

Before conducting the main Experiment 1, we had four pilot attempts which followed a slightly different preregistration (osf.io/f36su; in the same repository as the main study, osf.io/f49z7). The major differences to the main Experiment 1 were that we excluded participants when their accuracy dropped below 60% (rather than 50%) and that the outlier criterion was 2.5 Sn (rather than 3.0 Sn).

For the pilots, we preregistered to collect 48 participants and to replace participants whose data met specified exclusion criteria. Crucially, and as in the main experiment, we preregistered to collect small samples of eight participants each and to continue data collection only if the exclusion criteria were only met by a maximum of 2 participants (≤ 25%). Consequently, if a pilot had higher exclusion rates, we stopped data collection and adapted the study design to reduce exclusions. In sum, four pilots failed (see below for details on exclusions). Across the four pilot attempts, we adjusted the number of trials (Pilot 1: 1050, Pilot 2: 1170, Pilot 3: 1290, Pilot 4: 1860), the response deadline (Pilot 1: 700 ms, Pilot 2: 650 ms, Pilots 3 and 4: 675 ms), and the feedback duration (Pilots 1 and 2: 1000 ms, Pilots 3 and 4: 2000 ms). In Pilot 4, we excluded repetitions of irrelevant stimuli with the method of the main experiment. Specifically, the first element of each sequence of irrelevant stimuli was constrained to differ from the last two (one in Pilots 1-3) elements of the previous sequence of irrelevant stimuli. Furthermore, we excluded target repetitions in two successive trials and invited participants to two (one in Pilots 1-3) experimental sessions. In contrast to the main study, all participants were invited to the second session in the fourth pilot, irrespective of their error rate in the first session.

As in the main study, we had to exclude spurious force measurements from the unprocessed force data when all three force keys sent a value of 0 a.u. at the same time (Pilot 1: 0.3%, Pilot 2: 0.3%, Pilot 3: 0.3%, Pilot 4: 0.1%). Force was sampled effectively about every 4 ms (Pilot 1: M = 4.2 ms, SD = 0.02 ms, Pilot 2: M = 4.2 ms, SD = 0.02 ms, Pilot 3: M = 4.3 ms, SD = 0.05 ms, Pilot 4: M = 4.4 ms, SD = 0.06 ms). All further data treatment went as planned in the preregistration.

We excluded the practice blocks and participants who responded correctly in less than 60% of the remaining trials (Pilot 1: nobody, Pilot 2: two participants, Pilot 3: three participants, Pilot 4: two participants). As Pilot 3 already exceeded our maximum exclusion criterion of two participants, its data were not further analyzed. For the other pilots, we excluded the first and last trial of each block. We then excluded trials with an error in the preceding trial n-1 (Pilot 1: 22.3%, Pilot 2: 25.8%, Pilot 4: 23.0%). We

excluded early responses during fixation (all three remaining pilots < 0.1%), omission errors (Pilot 1: 8.2%, Pilot 2: 13.3%, Pilot 4: 11.0%), and additional responses during the blank (Pilot 1: 3.2%, Pilot 2: 1.6%, Pilot 4: 2.8%). We excluded target repetitions from trial n to trial n+1 in Pilot 1 (14.7%) and Pilot 2 (15.9%; Pilot 4 excluded target repetitions by design). We then excluded trials with errors in the upcoming trial n+1, that is, early responses during fixation (all three remaining pilots < 0.1%), omission errors (Pilot 1: 8.5%, Pilot 2: 14.2%, Pilot 4: 10.7%), and additional responses during the blank (Pilot 1: 3.3%, Pilot 2: 1.6%, Pilot 4: 2.7%). For this selection of trial sequences, we planned to conduct all analyses on PE_{n+1} and excluded commission errors for the analyses of all other dependent variables (Pilot 1: 9.4%, Pilot 2: 9.5%, Pilot 4: 7.1%). We excluded trials as outliers when any trial-specific value was greater than 2.5 Sn (Pilot 1: 27.3%, Pilot 2: 23.9%, Pilot 4: 25.3%). After these exclusions, six participants in Pilot 1, six participants in Pilot 2, and three participants in Pilot 4 had less than 10 observations in at least one of the design cells. Therefore, all four pilots exceeded our maximum exclusion criterion of two participants.

Appendix B

For Experiment 1, it might be argued that the observed differences in the execution of erroneous responses, relative to correct responses, are not indicative of cancelling the erroneous response but rather, are a byproduct of sub-threshold correction activity. In other words, the lower MF_{RD} and shorter RD for erroneous responses might be the result of increased response competition rather than pronounced cancellation. We conducted unplanned exploratory analyses where we determined the relationship between (i) PF on sub-threshold (i.e., not executed) responses and (ii) RD and MF_{RD} of above-threshold (i.e., executed) responses. Specifically, we added the PFs of the two sub-threshold responses in trial n (i.e., the two neutral responses in a correct trial or the correct and neutral response in an erroneous trial) and conducted a median split of these PF sums (≤ median vs. > median), separately for each participant and response accuracy (correct vs. erroneous) in trial n. We then analyzed RD and MF_{RD} in trial n with an analysis of variance (ANOVA) with two within-subject factors: (i) sub-threshold PF (low vs. high) and (ii) response accuracy in trial n (correct vs. erroneous). Significant interactions were investigated with follow-up, two-tailed paired-samples *t*-tests.

For RD, the main effect of the sub-threshold PF was not significant (low = 167 ms vs. high = 171 ms), $F(1, 46) = 2.06$, $p = .158$, $\eta_p^2 = .04$. There was, however, a main effect of response accuracy in trial n, as RD was shorter for erroneous trials than for correct trials (159 vs. 179 ms), $F(1, 46) = 19.41$, $p <$

.001, η_p^2 = .30 (see also H2). Finally, there was an interaction between these factors, $F(1, 46) = 10.65$, $p = .002$, $\eta_p^2 = .19$. In correct trials, RD was shorter when sub-threshold PF was low as compared to high (174 vs. 184 ms), $t(46) = -4.16$, $p < .001$, $d_z = -0.61$. In erroneous trials, however, the difference was not significant (158 vs. 160 ms), $t < 1$.

For MF_{RD}, there was a main effect of response accuracy in trial n as MF_{RD} was lower in erroneous trials than in correct trials (839 vs. 888 a.u.), $F(1, 46) = 11.37$, $p = .002$, $\eta_p^2 = .20$ (see also H2). However, neither the main effect of sub-threshold PF (low = 858 a.u. vs. high = 868 a.u.), $F(1, 46) = 1.38$, $p = .246$, $\eta_p^2 = .20$, nor the interaction (correct, low = 881 a.u. vs. correct, high = 894 a.u. vs. erroneous, low = 835 a.u. vs, erroneous, high = 842 a.u.), $F < 1$, were significant.

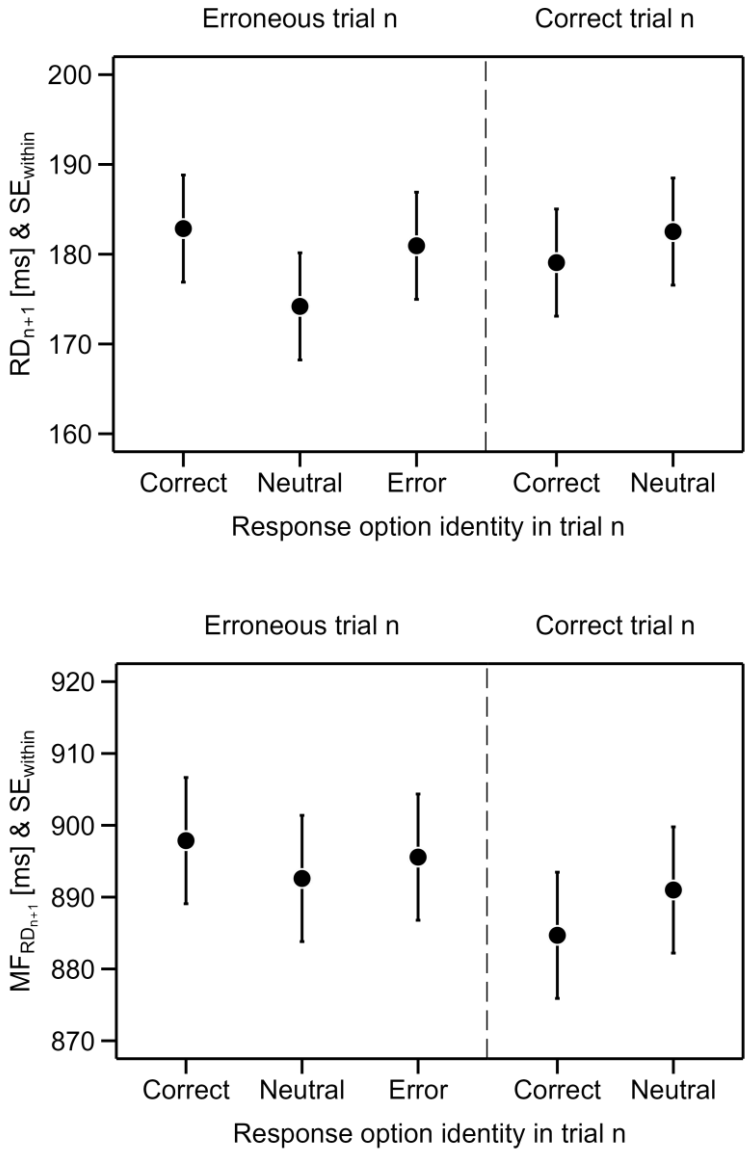
In sum, the exploratory analyses do not support the assumption that response competition is responsible for the observed differences between erroneous and correct responses in response execution.

Appendix C

In Experiment 1, we conducted separate ANOVAs on RD_{n+1} and $MF_{RD_{n+1}}$ with a single within-subjects factor: response option identity in trial n (correct vs. neutral vs. erroneous) after an erroneous trial n (see Fig. C1). Neither ANOVA yielded a significant main effect, both Fs < 1. We also assessed whether RD_{n+1} and $MF_{RD_{n+1}}$ for the neutral response from trial n differed following a correct trial n and an erroneous trial n. The two-tailed, paired-samples t-tests were not significant for RD_{n+1}, $t(44) = 1.42$, $p = .163$, $d_z = 0.21$, and $MF_{RD_{n+1}}$, $|t| < 1$.

Figure C1

Results for RD_{n+1} and MF_{RDn+1} in Experiment 1



Note. Response duration in trial n+1 (RD_{n+1}, top) and mean force in the response duration window of trial n+1 ($MF_{RD_{n+1}}$, bottom). The left side of the figure shows the correct, neutral, and erroneous response from an erroneous trial n. The right side of the figure shows the correct and neutral response from a correct trial n. Large, black dots indicate the mean. The error bars depict the within-subject standard error for the full factorial design (SE_{within}; Loftus & Mason, 1994).

Finally, in unplanned exploratory analyses, we comprehensively assessed differences in RD_{n+1} and $MF_{RD_{n+1}}$ between all combinations of response option identity and response accuracy in trial n (see Fig. C1). We conducted two separate ANOVAs with a single within-subjects factor that allowed for a comparison of all five combinations of response option identity and response accuracy (i.e., correct response from an erroneous trial, neutral response from an erroneous trial, erroneous response from

an erroneous trial, correct response from a correct trial, neutral response from a correct trial). There were no significant differences between the five conditions in RD_{n+1} or $MF_{RD_{n+1}}$, both Fs < 1.

Appendix D

As for RT_{n+1}, RD_{n+1}, and $MF_{RD_{n+1}}$ in Experiment 1, we assessed differences in $MF_{PreS_{n+1}}$ between all combinations of response option identity and response accuracy in trial n in unplanned, exploratory analyses (see Fig. 4). We conducted an ANOVA with a single within-subjects factor that allowed for a comparison of all five combinations of response option identity and response accuracy (i.e., correct response from an erroneous trial, neutral response from an erroneous trial, erroneous response from an erroneous trial, correct response from a correct trial, neutral response from a correct trial). Since $MF_{PreS_{n+1}}$ for the correct response from a correct trial n was not part of the design cells considered in our initial outlier detection, we discarded additional outliers for this analysis (1.4%).

We found a significant difference between the five conditions, $F(4, 184) = 7.10$, $p < .001$, $\eta_p^2 = .13$ (Greenhouse-Geisser corrected; $\varepsilon = .68$). Two-tailed, paired-samples t-tests between all conditions indicated that $MF_{PreS_{n+1}}$ for the correct response from a correct trial n was higher than for (a) the neutral response from a correct trial n and (b) each of the three responses from an erroneous trial n, all ts(46) ≥ 2.94, $p \leq .005$, $d_z \geq 0.43$. Each of the remaining six comparisons (i.e., between the neutral response from a correct trial n and each of the three responses from an erroneous trial n and between each of the three responses from an erroneous trial n) were not significant, all $|t|$s < 1.

In sum, the exploratory analyses revealed that only correct responses from trial n showed a pronounced subthreshold force in exploratory analyses, speculatively because these responses had the most forceful and longest execution in trial n. However, the pattern is not consistent with H4 or H5.

Appendix E

In Experiment 2, we conducted selected t-tests to test our hypotheses in RT_{n+1} and to conduct complementary secondary analyses in PE_{n+1}. For completeness, we report the remaining two-tailed paired samples t-tests here.

First, there is no reason to assume differences between the activation of the erroneous response and the correct response from an erroneous trial n. There was only a non-significant trend toward faster RT_{n+1}, $t(59) = 1.76$, $p = .083$, $d_z = 0.23$, but significantly higher PE_{n+1}, $t(59) = 2.67$, $p = .010$, $d_z = 0.34$,

for the erroneous than the correct response from an erroneous trial n. Second, as the results are in line with H3a, H4a, and H4b, RT_{n+1} should also be significantly faster for the correct response from a correct trial n than for the neutral response from an erroneous trial n, $t(59) = 9.85$, $p < .001$, $d_z = 1.27$. PE_{n+1} did not differ significantly between these conditions, $|t| < 1$. Third, the erroneous and the correct response from an erroneous trial n should be more active than the neutral response from a correct trial n. At the same time, response-general processes could slow down responding after erroneous trials n but not after correct trials n. These opposing influences could therefore cancel each other out and the absence of an effect would be difficult to interpret. Indeed, the neutral response from a correct trial n did not differ from both responses from an erroneous trial n in RT_{n+1} (correct: $t(59) = 1.48$, $p = .143$, $d_z = 0.19$; erroneous: $|t| < 1$). PE_{n+1} for the neutral response from a correct trial n was higher than for the correct response from an erroneous trial n, $t(59) = -2.19$, $p = .033$, $d_z = -0.28$, but did not differ significantly from the erroneous response from an erroneous trial n, $t(59) = 1.07$, $p = .290$, $d_z = 0.14$.

References

Adkins, T. J., Zhang, H., & Lee, T. G. (2024). People are more error-prone after committing an error. *Nature Communications*, *15*(1), 6422. https://doi.org/10.1038/s41467-024-50547-y

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., & Hocking, T. (2024). *data.table: Extension of 'data.frame'* [Computer software]. https://Rdatatable.gitlab.io/data.table, https://github.com/Rdatatable/data.table, https://r-datatable.com

Benini, E., Möller, M., Koch, I., Philipp, A. M., Qiu, R., & Mayr, S. (2024). Evidence of task-triggered retrieval of the previous response: A binding perspective on response-repetition benefits in task switching. *Psychonomic Bulletin & Review*, *31*(3), 1290–1300. https://doi.org/10.3758/s13423-023-02409-9

Bode, S., & Stahl, J. (2014). Predicting errors from patterns of event-related potentials preceding an overt response. *Biological Psychology*, *103*, 357–369. https://doi.org/10.1016/j.biopsycho.2014.10.002

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295x.108.3.624

Brewer, N., & Smith, G. A. (1989). Developmental changes in processing speed: Influence of speed-accuracy regulation. *Journal of Experimental Psychology: General*, *118*(3), 298–310. https://doi.org/10.1037/0096-3445.118.3.298

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454–465. https://doi.org/10.3758/s13414-011-0243-2

Fiehler, K., Ullsperger, M., & Cramon, D. Y. von (2004). Neural correlates of error detection and error correction: Is there a common neuroanatomical substrate? *The European Journal of Neuroscience*, *19*(11), 3081–3087. https://doi.org/10.1111/j.0953-816X.2004.03414.x

Fiehler, K., Ullsperger, M., & Cramon, D. Y. von (2005). Electrophysiological correlates of error correction. *Psychophysiology*, *42*(1), 72–82. https://doi.org/10.1111/j.1469-8986.2005.00265.x

Foerster, A., Moeller, B., Frings, C., & Pfister, R. (2023). What is left after an error? Towards a comprehensive account of goal-based binding and retrieval. *Attention, Perception, & Psychophysics*, *85*(1), 120–139. https://doi.org/10.3758/s13414-022-02609-w

Foerster, A., Moeller, B., Huffman, G., Kunde, W., Frings, C., & Pfister, R. (2022). The human cognitive system corrects traces of error commission on the fly. *Journal of Experimental Psychology: General*, *151*(6), 1419–1432. https://doi.org/10.1037/xge0001139

Foerster, A., Rothermund, K., Parmar, J. J., Moeller, B., Frings, C., & Pfister, R. (2021). Goal-based binding of irrelevant stimulus features for action slips. *Experimental Psychology*, *68*(4), 206–213. https://doi.org/10.1027/1618-3169/a000525

Foerster, A., Schaaf, M., Kunde, W., & Pfister, R. (2022). *Response activation and error processing.* https://doi.org/10.17605/OSF.IO/F49Z7

Foerster, A., Steinhauser, M., Schwarz, K. A., Kunde, W., & Pfister, R. (2022). Error cancellation. *Royal Society Open Science*, *9*(3), 210397. https://doi.org/10.1098/rsos.210397

Frings, C., Beste, C., Benini, E., Möller, M., Dignath, D., Giesen, C. G., Hommel, B., Kiesel, A., Koch, I., Kunde, W., Mayr, S., Mocke, V., Moeller, B., Münchau, A., Parmar, J., Pastötter, B., Pfister, R., Philipp, A. M., Qiu, R., . . . Schmalbrock, P. (2024). Consensus definitions of perception-action-integration in action control. *Communications Psychology*, *2*(1). https://doi.org/10.1038/s44271-023-00050-9

Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020). Binding and Retrieval in Action Control (BRAC). *Trends in Cognitive Sciences*, *24*(5), 375–387. https://doi.org/10.1016/j.tics.2020.02.004

Gjorgieva, E., & Egner, T. (2022). Learning from mistakes: Incidental encoding reveals a time-dependent enhancement of posterror target processing. *Journal of Experimental Psychology. General*, *151*(3), 718–730. https://doi.org/10.1037/xge0001105

Hochman, E. Y., Milman, V., & Tal, L. (2017). Evidence for aversive withdrawal response to own errors. *Acta Psychologica*, *180*, 147–154. https://doi.org/10.1016/j.actpsy.2017.09.007

Houtman, F., Castellar, E. N., & Notebaert, W. (2012). Orienting to errors with and without immediate feedback. *Journal of Cognitive Psychology*, *24*(3), 278–285. https://doi.org/10.1080/20445911.2011.617301

Houtman, F., & Notebaert, W. (2013). Blinded by an error. *Cognition*, *128*(2), 228–236. https://doi.org/10.1016/j.cognition.2013.04.003

Jentzsch, I., & Dudschig, C. (2009). Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Quarterly Journal of Experimental Psychology*, *62*(2), 209–218. https://doi.org/10.1080/17470210802240655

Jentzsch, I., & Leuthold, H. (2006). Control over speeded actions: A common processing locus for micro- and macro-trade-offs? *Quarterly Journal of Experimental Psychology (2006)*, *59*(8), 1329–1337. https://doi.org/10.1080/17470210600674394

Jones, P. R. (2019). *RousseeuwCrouxSn.m* (Version 22/01/2019) [Computer software]. https://github.com/petejonze/psychosim/blob/9b477fba359035dbd46594994d6cedf5eacc0f13/RousseeuwCrouxSn.m

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological Review*, *97*(2), 253–270. https://doi.org/10.1037/0033-295X.97.2.253

Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica*, *43*(3), 199–224. https://doi.org/10.1016/0001-6918(79)90026-X

Lawrence, M. A. (2016). *ez: Easy analysis and visualization of factorial* [Computer software].

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. https://doi.org/10.3758/BF03210951

Mohr, H., Zwosta, K., Markovic, D., Bitzer, S., Wolfensteller, U., & Ruge, H. (2018). Deterministic response strategies in a trial-and-error learning task. *PLoS Computational Biology*, *14*(11), e1006621. https://doi.org/10.1371/journal.pcbi.1006621

Moretti, L., Koch, I., Steinhauser, M., & Schuch, S. (2021). Errors in task switching: Investigating error aftereffects in a N-2 repetition cost paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(10), 1720–1737. https://doi.org/10.1037/xlm0001034

Moretti, L., Koch, I., Steinhauser, M., & Schuch, S. (2023). Disentangling task-selection failures from task-execution failures in task switching: An assessment of different paradigms. *Psychological Research*, *87*(3), 929–950. https://doi.org/10.1007/s00426-022-01708-5

Notebaert, W., Houtman, F., van Opstal, F., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition*, *111*(2), 275–279. https://doi.org/10.1016/j.cognition.2009.02.002

Notebaert, W., & Verguts, T. (2011). Conflict and error adaptation in the Simon task. *Acta Psychologica*, *136*(2), 212–216. https://doi.org/10.1016/j.actpsy.2010.05.006

Núñez Castellar, E., Kühn, S., Fias, W., & Notebaert, W. (2010). Outcome expectancy and not accuracy determines posterror slowing: Erp support. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 270–278. https://doi.org/10.3758/CABN.10.2.270

Parmar, J. J., Foerster, A., Pfister, R., & Rothermund, K. (2022). Frankly, my error, I don't give a damn: Retrieval of goal-based but not coactivation-based bindings after erroneous responses. *Journal of Cognition*, *5*(1), 34. https://doi.org/10.5334/joc.224

Pedersen, T. L. (2023). *patchwork: The Composer of Plots* [Computer software]. https://CRAN.R-project.org/package=patchwork

Pfister, R., Bogon, J., Foerster, A., Kunde, W., & Moeller, B. (2022). Binding and retrieval of response durations: Subtle evidence for episodic processing of continuous movement features. *Journal of Cognition*, *5*(1), 23. https://doi.org/10.5334/joc.212

Pfister, R., & Janczyk, M. (2016). schoRsch: An R package for analyzing and reporting factorial experiments. *The Quantitative Methods for Psychology*, *12*(2), 147–151. https://doi.org/10.20982/tqmp.12.2.p147

Pfister, R., Neszmélyi, B., & Kunde, W. (2023). Response durations: A flexible, no-cost tool for psychological science. *Current Directions in Psychological Science, 32*(2), 160-166. https://doi.org/10.1177/0963721422114169

Pfister, R., Foerster, A., Schwarz, K. A., Varga, S., Steinhauser, M., & Kunde, W. (2025). Error-related brain activity indicates immediate auto-cancellation of action slips. *Psychophysiology 62*(10), e70160. https://doi.org/10.1111/psyp.70160

Pfister, R., Tonn, S., Schaaf, M., & Wirth, R. (2024). mousetRajectory: Mouse tracking analyses for behavioral scientists. *The Quantitative Methods for Psychology*, *20*(3), 217–229. https://doi.org/10.20982/tqmp.20.3.p217

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Computer software]. https://www.R-project.org/

Rabbitt, P. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*(2), 264–272. https://doi.org/10.1037/h0022853

Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology*, *70*(2), 305–311. https://doi.org/10.1111/j.2044-8295.1979.tb01687.x

Rabbitt, P. (2002). Consciousness is slower than you think. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *55*(4), 1081–1092. https://doi.org/10.1080/02724980244000080

Reason, J. (1990). *Human error*. Cambridge University Press.

Reason, J. (1995). Understanding adverse events: Human factors. *Quality in Health Care*, *4*(2), 80–89. https://doi.org/10.1136/qshc.4.2.80

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283. https://doi.org/10.1080/01621459.1993.10476408

Schaaf, M., Kunde, W., & Wirth, R. (2022). Evidence for initially independent monitoring of responses and response effects. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(2), 128–138. https://doi.org/10.1037/xhp0000979

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *The Journal of Applied Psychology*, *51*(3), 300–304. https://doi.org/10.1037/h0020586

Steinborn, M. B., Flehmig, H. C., Bratzke, D., & Schröter, H. (2012). Error reactivity in self-paced performance: Highly-accurate individuals exhibit largest post-error slowing. *Quarterly Journal of Experimental Psychology (2006)*, *65*(4), 624–631. https://doi.org/10.1080/17470218.2012.660962

Steinhauser, M., Ernst, B., & Ibald, K. W. (2017). Isolating component processes of posterror slowing with the psychological refractory period paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 653–659. https://doi.org/10.1037/xlm0000329

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*(3), 517–534. https://doi.org/10.1037/0096-1523.32.3.517

Steinhauser, M., & Hübner, R. (2008). How task errors affect subsequent behavior: Evidence from distributional analyses of task-switching effects. *Memory & Cognition*, *36*(5), 979–990. https://doi.org/10.3758/MC.36.5.979

Steinhauser, R., & Steinhauser, M. (2021). Adaptive rescheduling of error monitoring in multitasking. *NeuroImage*, *232*, 117888. https://doi.org/10.1016/j.neuroimage.2021.117888

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tonn, S., Schaaf, M., Kunde, W., & Pfister, R. (2023). Action representations in prevention behavior: Evidence from motor execution. *Cognition*, *234*, Article 105370. https://doi.org/10.1016/j.cognition.2023.105370

Weissman, D. H. (2019). Let your fingers do the walking: Finger force distinguishes competing accounts of the congruency sequence effect. *Psychonomic Bulletin & Review*, *26*(5), 1619–1626. https://doi.org/10.3758/s13423-019-01626-5

Weissman, D. H., & Schmidt, J. R. (2024). Proactive response preparation contributes to contingency learning: Novel evidence from force-sensitive keyboards. *Psychological Research*, *88*(4), 1182–1202. https://doi.org/10.1007/s00426-024-01940-1

Weissman, D. H., Schmidt, J. R., & Spinelli, G. (2024). Modulations of response activation contribute to block-wide control: Evidence from proportion congruency effects in the prime-probe task. *Journal of Experimental Psychology: Learning, Memory, & Cognition.* Advance online publication. https://doi.org/10.1037/xlm0001404

Welford, A. T. (1959). Evidence of a Single-Channel Decision Mechanism Limiting Performance in a Serial Reaction Task. *Quarterly Journal of Experimental Psychology*, *11*(4), 193–210. https://doi.org/10.1080/17470215908416312

Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology*, *55*(3). https://doi.org/10.1111/psyp.13041

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wirth, R., Janczyk, M., & Kunde, W. (2018). Effect monitoring in dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 553–571. https://doi.org/10.1037/xlm0000474

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*(4), 931–959. https://doi.org/10.1037/0033-295x.111.4.939