The Human Cognitive System Corrects Traces of Error Commission on the Fly

Anna Foerster1, Birte Moeller2, Greg Huffman3, Wilfried Kunde1, Christian Frings2, and Roland Pfister1

1 Department of Psychology III, University of Würzburg
2 Department of Cognitive Psychology, University of Trier
3 Leidos, Inc., San Diego, California, United States

Human perception and action rely on a fundamental binding mechanism that forges integrated event representations from distributed features. Encountering any one of these features later on can retrieve the whole event, thus expediting cognitive processing. The traditional view on binding confines it to successful action episodes, holding that the human cognitive system does not leverage errors for optimizing corresponding event representations. Here we use sequential analyses of erroneous action episodes to explore whether binding promotes future successful behavior even when actions go awry. Results indicate that the processes leading to binding integrate different aspects of the action episode in a highly efficient and flexible manner to privilege future correct actions and prepare the ground for error-based learning.

Keywords: action control, error processing, binding and retrieval

Every error comes with an opportunity to learn from one’s mistakes, and error-based learning provides powerful boosts to future performance, decision making and memory (Metcalfe, 2017). Error detection further yields an immediate reinforcement learning signal and adapts processing to avoid errors in the future (Botvinick et al., 2001; Holroyd & Coles, 2002). Such cognitive effects of error commission are accompanied by error-based adaptation in motor skill learning in order to update and correct internal models for action and perception (Diedrichsen et al., 2005; Shadmehr et al., 2010). Against this background, it is perhaps surprising that the representation of erroneous actions itself has received relatively little interest. The present experiments therefore aimed at capturing the representation of erroneous responses from the basic perspective of feature binding in action and perception.

Specifically, agents compile integrated representations of sensorimotor action episodes by binding features of their actions and features of concurrent stimulation into event files (Frings et al., 2020; Hommel, 2004; Hommel et al., 2001; Kahneman & Treisman, 1984; Kahneman et al., 1992; Schumacher & Hazeline, 2016). Re-encountering individual features retrieves such event files as a whole, thereby creating an efficient short-cut for adaptive action control. Compiling action episodes for later retrieval is only adaptive, however, if it paves the way for future successful behavior. Indeed, current theories assume that only successful actions trigger event file compilation (Hommel, 2005). Here we challenge this basic and previously untested assumption by studying binding for action slips, that is, failures to act as intended. We demonstrate that binding is indeed an adaptive mechanism, but not by being switched off after action slips. Instead, binding operates in a highly selective manner that captures features of intended actions despite a strongly activated representation of erroneous actions, thereby correcting traces of error commission on the fly.

We propose three accounts of binding for action slips that make clear-cut predictions about whether and which action and stimulus features become integrated in the face of error commission. All three accounts agree with current theorizing on event files that these compounds are created whenever any stimuli appear in sufficiently close temporal proximity to an action (Hommel, 2005). Current theorizing further adopts a success-based account, assuming that binding is conditional on successful responding, thus predicting no binding for action slips. However, empirical data from research on binding and retrieval as well as on error processing suggests that binding might also be effective in erroneous action episodes. We therefore propose two competitor accounts that both assume binding for action slips but differ in the type of response that enters these bindings. The co-activation account holds that binding occurs by default after executing any action, which would...
result in binding of stimulus features to features of the erroneous action. Recent observations on binding for unsuccessful stopping of prepared actions support this account tentatively (Giesen & Rothermund, 2014). The goal-based account emerges from findings on error detection and subsequent correction responses (Crump & Logan, 2013; Gehring et al., 2012; Logan & Crump, 2010; Rabbitt, 1966, 1978). These findings suggest that a representation of the intended action is held active even during error commission, allowing for swift corrections after error commission. Holding this representation active, however, may also allow for binding the intended, but not actually executed, action plan to features of the current stimulation.1 Crucially, both binding of the intended correct response and binding of the executed erroneous responses can be adaptive, depending on what aspects of the situation are included in an event file. When integrating actions with stimuli that indicate the correct response (i.e., task-relevant stimulation), goal-based binding would result in maximally adaptive bindings. When integrating actions with their effects on the environment, however, binding through co-activation actually might prove efficient. Therefore, Experiment 1 investigated binding of task-relevant stimuli to responses whereas Experiment 2 expanded the focus to binding of erroneous responses to their effects in the agent’s environment.

Experiment 1

We investigated binding for action slips by means of sequential analyses of a speeded choice reaction time task that we optimized to provoke errors from the participants (see Figure 1). Critically, the task mapped four target stimuli to two response options so that we could assess whether repeating the target stimulus would retrieve the participant’s previous response (Frings et al., 2020). We assessed trial sequences with a correct response in the current trial and a commission error or correct response in the preceding trial. Comparing trial sequences with target repetitions to target changes that called for the same correct response after an error had been committed, provides a measure of binding and retrieval with clear predictions for all three accounts (Figure 2A; see Frings et al., 2020; Henson et al., 2014; Hommel et al., 2004; Moeller, Frings, et al., 2016). The success-based account assumes binding not to occur in the case of action slips, thus predicting similar performance for correct response repetitions after errors, irrespective of whether targets repeat or change. The co-activation account assumes that a target repetition retrieves the previously bound erroneous response, thus predicting performance decrements for correct response repetitions after errors when the target repeats rather than changes. Finally, the goal-based account assumes that target repetitions retrieve bindings with the previously intended, correct response, leading to performance boosts for correct response repetitions if the target repeats rather than changes.

Method

This research complies with the ethical regulations of the Ethics Committee of the local Institute of Psychology, the German Psychological Society, and the German Research Foundation.

Participants

Binding effects for task-relevant stimuli and correct responses are typically reported to be of large size, for example, $d_z = 2.70$ (Exp. 1b in Hommel, 1998) or $d_z = 1.15$ (Exp. 1a in Frings et al., 2007). Using the smaller estimate of $d_z = 1.15$ would require 16 participants for a power of 99%. For the analysis of binding effects for incorrect responses, we anticipated fewer observations in each design cell, however, which might reduce effect sizes to an unknown degree. We therefore decided to triple this sample size and preregistered to collect a pilot sample of eight participants to assess whether the experimental design elicits a reasonable amount of errors, followed by a new sample of 48 participants (https://aspredicted.org/vm5pe.pdf). We only report data from the latter sample here (33 female; 46 right-handers; age: $M = 29$ years, $SD = 11$ years). The chosen sample size comes with a power of 95% for effects of less than half the size as for correct responses ($d_z = 0.53$) and a power of 80% to detect effects of at least $d_z = 0.41$, assuming an alpha of 5% in a two-tailed test (calculated with the power.t.test function in the statistics package R version 4.0.2). Even after participant exclusions, the design came with a power of 95% for $d_z = 0.56$ and 80% for $d_z = 0.43$ (see Data treatment and analyses for details on data exclusion). All participants signed a written informed consent sheet and received course credit or monetary compensation for their participation.

Apparatus and Stimuli

The experiment took place in sessions with up to five participants taking part in parallel, separated by dividers. Each workstation provided a 24” screen (display resolution; $1920 \times 1080$ pixels; refresh rate: 100 Hz) and a standard German QWERTZ keyboard. To exclude potential confounds by perceptual repetition effects, we opted for simple alphanumeric stimuli which have been shown to prevent perceptual priming across experimental trials (Pashler & Baylis, 1991). Participants used the keys $F$ and $J$ to categorize the target letters $R$ and $N$ with one key and $V$ and $K$ with the other key using their index fingers. The assignment of pairs of target letters to the two keys was counterbalanced across participants. Targets appeared centrally in white font against a black background. Above, below, left and right of the target appeared an irrelevant letter in white font that was drawn from a pool of eight letters (i.e., $O$, $W$, $X$, $U$, $Z$, $L$, $H$, $A$). We included these irrelevant letters to increase difficulty of the task and commission errors through perceptual noise. The same letter was displayed at all four positions (horizontal displacement: 58 px, vertical displacement: 65 px, relative to the center of the target letter).

Procedure

Participants were instructed about the mapping rule and instructions further encouraged participants to ignore all task-irrelevant stimuli. We prompted participants to respond as fast as and accurately as possible throughout the experiment. Participants learned that they would first work through a practice block where they

---

1 Albeit detectable, the activation of the correct response is relatively weak in case of action slips, as these events evidently tend to arise from a lack of goal-directed, executive processing (e.g., Hoffmann & Beste, 2015).
would receive feedback for each of their responses. Before starting the practice block, they were given the opportunity to go through all instructions again if they wanted to.

Figure 1A shows a schematic of the trial procedure. Each trial started with a white fixation cross for 750 ms, after which the stimuli appeared and stayed on screen until response or for a maximum of 600 ms. In the practice block, participants received feedback for 1000 ms in each trial: “Good!” (German original: “Gut!”) in green font color for correct responses, “Too slow!” (German original: “Zu langsam!”) in red font color if no response had been delivered within 600 ms (omission error), and “Wrong!” (German original: “Falsch!”) in red font if participants delivered a response with the left key when the right key was correct and vice versa (commission error) or responded with any other than the instructed keys or delivered multiple responses (miscellaneous error). After the practice block, participants received the instruction that in all following experimental blocks, they would only receive feedback for omission errors and were reminded of that at the beginning of each block. This established equivalent procedures in trials with correct responses and commission errors. After the practice block, participants received feedback about how fast they had responded in the preceding block, how many errors they had committed (sum

Note. (A) Participants had to respond to the central target letter within 600 ms and they had to ignore the surrounding letters. Two target letters were mapped to one response whereas another pair of target letters was mapped to the other response (V, K vs. R, N). The depicted trial sequence introduces a target change and a repetition of the correct response as both targets map to the same response key. (B) On the left, the target and the correct response repeat in successive trials, whereas in the middle the target changes although the correct response repeats. The comparison of these two condition sequences allows for assessing binding effects after correct and erroneous responses. The third condition sequence yields a target and response change. See the online article for the color version of this figure.
of commission and miscellaneous errors) and how often they had failed to respond within the response deadline of 600 ms.

Participants went through 20 blocks of 56 trials each (one practice, the remaining experimental). The array of eight irrelevant stimuli was randomized at the beginning of the experiment and after each eighth trial while ensuring that successive trials never featured the same irrelevant letter to avoid response retrieval from these additional stimuli. The target was chosen randomly in the first trial of each block. For all following trials, we implemented a pseudorandom strategy to ensure that the design resulted in sufficiently many observations for all condition sequences in the case of errors (i.e., target repetition | correct response repetition, target change | correct response repetition and target change | correct response change; see Figure 1B). We therefore drew the current condition sequence from an array that featured each of the three conditions three times. We randomized this array after each ninth commission error or after the registration of nine of the other responses (correct, omission, random error). After we had determined the current condition sequence with this procedure, the target was determined randomly while considering whether the target or response had to repeat or change.

We had originally planned to use separate arrays for commission errors and all other responses (we implemented only one due to an initial programming error), though the chosen procedure resulted in an optimal balance of conditions. Across our whole sample of 48 participants, trial sequences with a correct response in the preceding trial featured 32.9% target repetitions | correct response repetitions, 33.3% target changes | correct response repetitions and target changes | correct response repetition and trial sequences with a commission error in the preceding trial amounted to 33.3% target repetitions | correct response repetitions, 33.4% target changes | correct response repetitions and 33.3% target changes | correct response repetitions.

Data Treatment

We discarded the first trial of each block because our analyses targeted performance as a function of the preceding trial. We then discarded trials with miscellaneous errors (3.5%) or omission errors (4.8%) in the preceding trial. We computed the percentage of omission errors (number of omission errors/sum of the number of correct trials and omission errors) and commission errors (number of commission errors/sum of the number of correct trials and commission errors) for this subset of the data. For all other dependent variables, we also excluded erroneous trials (17.9%). We identified trials as outliers and excluded them if their response time deviated more than 2.5 standard deviations (SDs) from their respective cell mean (1.2%). Four participants came with less than ten observations in at least one of the experimental cells after preprocessing and thus had to be excluded from all statistical analyses as per our preregistration (see Figure 1B for a schematic of the different conditions sequences that applied for preceding correct responses and preceding errors). The remaining 44 participants delivered on average 257 usable trials ($SD = 39$) in target repetition | correct response repetition sequences, 227 trials ($SD = 40$) in target changes | correct response repetitions sequences, and 197 trials ($SD = 55$) in target change | response change sequences after a correct response in the preceding trial. After a commission error in the preceding trial, they delivered on average 33 trials ($SD = 12$) in target repetition | correct response repetition sequences, 31 trials ($SD = 11$) in target changes | correct response repetitions sequences and 29 trials ($SD = 10$) in target change | response change sequences.

Data Analysis

We analyzed response times in a $2 \times 3$ repeated-measures analysis of variance (ANOVA) with the factors preceding accuracy (correct vs. commission error) and condition sequence (target repetition | correct response repetition vs. target change | correct response repetition and target change | correct response changes). In case of a significant main effect of condition sequence and a nonsignificant interaction of both factors in the omnibus ANOVA, binding effects were followed up by a two-tailed paired-samples $t$-test comparing target repetition | correct response repetition and target change | correct response repetitions vs. target change | correct response repetition sequences averaged over preceding accuracy. In case of a significant interaction of both factors in the omnibus ANOVA, binding effects were scrutinized in a reduced $2 \times 2$ ANOVA excluding target change | correct response change sequences. An interaction in this reduced ANOVA was further analyzed in two-tailed paired-samples $t$-tests to test for binding effects separately for preceding correct and erroneous responses. For all $t$-tests, we report the effect size $d$, $\left( \frac{M_{ART} - M_{SDRT}}{SD_{ART}} \right)$. We also employed these analyses for the percentage of commission errors and omission errors in the current trial, to test for potential speed-accuracy trade-offs. We further repeated these analyses for the relative variance of response times ($RV_{4T};$ $\frac{SD_{RT} + SD_{ART}}{M_{RT}}$) and the difference in response durations between successive trials ($\Delta RD$), whereby smaller values would be indicative of a retrieval of the temporal features of the previously bound response (see also Bogon et al., 2017; for binding of the temporal features of a stimulus). For factors with more than two levels, we tested for violations of sphericity and report Greenhouse-Geisser corrections along with the corresponding $\varepsilon$ estimate when necessary.

We further employed exploratory analyses to gain in-depth information on potential sources for differences in binding effects following correct and erroneous responses. For one, we calculated Pearson-correlations of target-binding effects for correct and erroneous episodes in response times, percentages of commission and omission errors (two-tailed testing). Second, we calculated bimodality coefficients and skewness parameters for the individual response time data of each participant for trial sequences with repetitions of the target and the correct response, separately for

---

2 The main effect of preceding accuracy captures post-error slowing as assessed via the traditional comparison of correct post-error and correct post-correct trials. We verified that this computation resulted in an unbiased estimate of post-error slowing for the present data (Pfister & Foerster, 2021).

3 In the present design, changes of the correct response are mainly implemented to allow for a random trial sequence, and we only included it in the initial analyses to provide a comprehensive analysis of the data at hand. As in previous studies, we expected this condition to come with slow responses as compared to correct response repetitions (e.g., Moeller, Frings, et al., 2016). This pattern is not predicted from a perspective on binding and retrieval, but it might best be explained in terms of subjective expectations of either response repetition or response change trials in such designs (Soetens et al., 1985).
Results

Figure 2 shows the predictions of the three theoretical accounts alongside the main results of Experiment 1 (see Table A1 in the Appendix for a full overview of the descriptive data for each dependent variable and experimental cell).

Response Times

Figure 2B shows response times as a function of preceding accuracy and condition sequence. Responses were slower after a commission error than after a correct response, $F(1, 43) = 36.47$, $p < .001$, $\eta^2 = .46$, and response times differed between condition sequences, $F(2, 86) = 91.36$, $p < .001$, $\eta^2 = .68$, with particularly
fast responses for target repetitions | correct response repetitions and slow responses for target changes | correct response changes. The two factors entered a sizable interaction, $F(2, 86) = 45.63$, $p < .001$, $\eta^2_g = .52$ ($\epsilon = .70$).

The reduced analysis (excluding target change | correct response change sequences) showed that binding effects between correct and erroneous action episodes differed in strength, as indicated by a significant two-way interaction of preceding accuracy and condition sequence, $F(1, 43) = 37.06$, $p < .001$, $\eta^2_g = .46$. This analysis further revealed substantial post-error slowing, $F(1, 43) = 79.74$, $p < .001$, $\eta^2_g = .65$, as well as overall faster responses for target repetitions than for target changes, $F(1, 43) = 47.95$, $p < .001$, $\eta^2_g = .53$. For sequences of two successive correct responses, repetitions of the correct response were faster with target repetitions than with target changes ($M = 34 ms, SD = 24 ms$), $t(43) = 9.37$, $p < .001$, $d_z = 1.41$, 95% CI [27 ms, 41 ms], and, crucially, the same was true for responses following commission errors ($M = 12 ms, SD = 26 ms$), $t(43) = 2.95$, $p = .005$, $d_z = 0.44$, 95% CI [4 ms, 19 ms].

**Commission Errors**

Preceding accuracy did not affect the percentage of commission errors significantly (see Figure A1A in the Appendix), $F(1, 43) = 3.41$, $p = .072$, $\eta^2_g = .07$, but the main effect of condition sequence was significant, $F(2, 86) = 48.39$, $p < .001$, $\eta^2_g = .53$ ($\epsilon = .66$), indicating that the percentage of commission errors was lowest for target repetitions | correct response repetitions and highest for target changes | correct response changes. The interaction of the two factors was also significant, $F(2, 86) = 29.47$, $p < .001$, $\eta^2_g = .41$ ($\epsilon = .82$). The follow-up 2 × 2 ANOVA also showed a significant main effect of preceding accuracy, $F(1, 43) = 27.78$, $p < .001$, $\eta^2_g = .39$, as well as more accurate response repetitions for target repetitions than target changes, $F(1, 43) = 59.73$, $p < .001$, $\eta^2_g = .58$. The interaction was also significant, $F(1, 43) = 8.34$, $p = .006$, $\eta^2_g = .16$. The binding effect was larger after correct responses ($M = 8.88%, SD = 4.88%$), $t(43) = 12.16$, $p < .001$, $d_z = 1.83$, 95% CI [7.3%, 10.3%], than after commission errors ($M = 3.6%, SD = 10.2%$), $t(43) = 2.33$, $p = .024$, $d_z = 0.35$, 95% CI [0.5%, 6.7%], but clearly evident after both types of responses.

**Omission Errors**

Omission errors emerged more frequently after commission errors than after correct responses in the preceding trial (see Figure A1B in the Appendix), $F(1, 43) = 37.16$, $p < .001$, $\eta^2_g = .46$. Furthermore, they varied with condition sequence, $F(2, 86) = 23.55$, $p < .001$, $\eta^2_g = .35$, whereas the interaction of both factors was not significant, $F(2, 86) = 1.23$, $p = .296$, $\eta^2_g = .03$. The percentage of omission errors was lowest for target repetitions | correct response repetitions and highest for target changes | correct response changes, with a significant binding effect ($M = 3.4%, SD = 3.9%$), $t(43) = 5.74$, $p < .001$, $d_z = 0.87$, 95% CI [2.2%, 4.6%].

**Variance of Response Times**

The RV$_{RT}$ did not differ regarding preceding accuracy, $F(1, 43) = 2.51$, $p = .121$, $\eta^2_g = .06$. However, there was a significant main effect of condition sequence, $F(2, 86) = 9.79$, $p = .001$, $\eta^2_g = .19$ ($\epsilon = .64$), indicating that the RV$_{RT}$ was lowest for target changes | correct response changes, followed by target repetitions | correct response repetitions and finally highest for target changes | correct response changes. The binding effect was significant ($M = 3 ms, SD = 4 ms$), $t(43) = 4.31$, $p < .001$, $d_z = 0.65$, 95% CI [1 ms, 4 ms]. There was a nonsignificant, descriptive trend toward an interaction of both factors, $F(2, 86) = 3.04$, $p = .078$, $\eta^2_g = .07$ ($\epsilon = .62$).

**Response Durations**

ARDs were larger after commission errors than after correct responses, $F(1, 43) = 58.92$, $p < .001$, $\eta^2_g = .58$, whereas condition sequence did not exert a significant impact, $F(2, 86) = 1.89$, $p = .158$, $\eta^2_g = .04$. The two factors interacted, $F(2, 86) = 58.01$, $p < .001$, $\eta^2_g = .57$ ($\epsilon = .74$). In the follow-up 2 × 2 ANOVA (excluding target change | correct response change sequences), ARDs were again larger in post-error trials, $F(1, 43) = 91.46$, $p < .001$, $\eta^2_g = .68$, whereas the main effect of condition sequence, $F(1, 43) = 2.77$, $p = .103$, $\eta^2_g = .06$, and the interaction were not significant, $F(1, 43) < 1$.

**Distributional Analyses**

Larger binding effects for correct episodes coincided with larger binding effects for erroneous episodes across participants in response times (Figure 2C), $r = .52$, $t(42) = 3.90$, $p < .001$. Both distributions of binding effects in response times were unimodal (bimodality coefficients ≤ .265). Target-binding effects after a preceding correct and a preceding erroneous response did not correlate in percentages of commission errors, $r = - .15$, $t(42) = 0.98$, $p = .333$, or percentages of omission errors (correct: $M = 2.5%, SD = 6.9%$; error: $M = 4.3%, SD = 3.7%$), $r = .03$, $t(42) = 0.18$, $p = .855$. The bimodality coefficient was similar after commission errors ($M = 0.19$, $SD = 0.08$) and after correct responses ($M = 0.42$, $SD = 0.08$; see Figure A2B in the Appendix), $t(43) = 0.22$, $p = .824$, $d_z = 0.03$, 95% CI [-.028, .034]. The skewness after commission errors ($M = 0.43$, $SD = 0.39$) was significantly less positive than after correct responses ($M = 0.41$, $SD = 0.32$), $t(43) = 5.36$, $p < .001$, $d_z = 0.81$, 95% CI [.230, .507]. The Pearson-correlations between post-error slowing ($M = 63 ms, SD = 69 ms$) and binding in response times were neither significant in erroneous trial sequences, $r = - .22$, $t(42) = 1.48$, $p = .147$, nor in correct trial sequences, $r = .12$, $t(42) = 0.75$, $p = .458$.

**Discussion**

The first experiment yielded binding effects in response times and error rates not only for correct responses but also for action slips. That is, repeating a correct response was easier for target repetitions than for target changes and this was true also if an erroneous response had occurred in the preceding trial. Further, corresponding response times were also less variable for target repetitions than for target changes. Together, these results indicate that action slips do not only come with a co-activation of the correct response as suggested in preceding studies (Crump & Logan, 2013; Gehring et al., 2012; Logan & Crump, 2010; Rabbitt, 1966; 1978). They also provoke bindings between the correct response and stimulus features that can be retrieved in a subsequent action episode as proposed in the goal-based account. That is, binding corrects traces of error commission on the fly, and this correction of erroneous representations sets the stage for future successful
behavior by representing a past error mainly in terms of its underlying goals.

Observing this pattern is especially remarkable when considering that executed responses come with actual sensory (e.g., tactile) feedback that can be integrated into an event file (e.g., Friedrich et al., 2020; Pfister, 2019), whereas such feedback can only be predicted for merely intended actions. This imbalance likely results in richer event files for correct responses that can be retrieved more readily as mirrored in the stronger binding effects for correct responses as compared to action slips. This interpretation appears especially plausible here because the distributional analyses did not point to a mixture of binding mechanisms for erroneous responses. The medium-sized positive correlation between binding effects after erroneous and correct responses in response times points to shared aspects of binding, which might relate to binding of abstract features (e.g., “right response”). Differences in response durations of successive responses might be a good correlate for binding of execution-specific features. That is, smaller binding effects should emerge for erroneous compared to correct action episodes, however, binding effects did not even emerge for correct action episodes in this measure. Continuous, and task-irrelevant features of a response, such as its duration might therefore not be incorporated in stimulus-response bindings at all, though this conclusion certainly awaits further empirical clarification. Future experiments might further differentiate between features of the response that are abstract and equally available for both, the intended and the executed response, and features that become fully available only with the actual execution of a response. Binding effects might only differ for the latter feature category but not for abstract features.

The former considerations focused on the type of features that enter bindings if agents respond correctly or commit an error. The complementary perspective is whether persistent error processing affects retrieval in an upcoming action episode. Although the data showed typical post-error slowing effects, these effects did not correlate significantly with binding effects after erroneous or after correct action episodes. These absent correlations tentatively indicate that retrieval does not contribute to post-error slowing and rather happens seamlessly after correct and erroneous responses alike or that post-error slowing does not hamper retrieval processes. A third highly interesting perspective would be whether ongoing processing of a previous commission error changes binding of a current response, but this question goes beyond the scope of the current study.

Experiment 2

Although goal-based binding provides an adaptive take on stimulus-response compounds, such binding would be detrimental if it also applied to responses and their resulting effects. That would imply that agents bind the intended correct response to the observed effect although this response did not produce the effect. The effects of an erroneous action are still caused by that performed action, irrespective of how this action aligns with the agent’s intentions. Although action effects in most binding and retrieval designs do not share contingencies with participants’ responses, binding of responses to their effects might still be adaptive in that it allows for the detection of potential contingencies in the first place (Huffman & Brockmole, 2020). From that perspective, linking erroneous actions to their effects provides an agent with a powerful means to learn by errors (Mettea, 2017). A maximally adaptive system would therefore rely on binding through co-activation rather than goal-based binding of actions and their effects. Whether or not the human cognitive system is actually structured in such a way is an open question, however. Such a potential dissociation of different binding mechanisms—that is, goal-based binding of targets and co-activation of effects—is backed up by the finding that features of the response enter binary bindings with stimulus and effect features, respectively, without evidence for bindings between the latter two (Moeller et al., 2019).

We tested this hypothesis in a second experiment that employed common methodology for investigating short-term binding between responses and following sensory effects (Dutzi & Hommel, 2009; Moeller, Pfister, et al., 2016). Experiment 2 thus employed a prime-probe design on each trial in which effects of the prime response are presented as stimuli in the probe (see Figure 3). Participants gave speeded responses to a numeric target in the prime, and the prime response randomly produced one of two noncontingent sound effects. The probe commenced by a sound that could either repeat or change relative to the prime effect, and participants performed a free-choice response by either repeating the prime response or by performing the alternative response. Again, the three accounts make clear and contrasting predictions on how the sound sequence should affect the frequency of prime response repetitions in the probe (Figure 4A). The success-based account assumes no binding and therefore predicts no differences between sound sequences. The co-activation account proposes binding of the erroneous response and its effect, predicting more response repetitions from prime to probe if the sound also repeats from prime to probe than if it changes. In contrast, the goal-based account assumes binding between the intended correct response and the experienced effect, predicting fewer response repetitions for sound repetitions than changes.

Method

Participants

Following our previous power calculations and the experiences from Experiment 1, we gathered data from 48 new participants (preregistration: https://osf.io/fzq2v; 41 female; 43 right-handers; age: M = 26 years, SD = 11 years). All participants signed a written informed consent sheet and received course credit or monetary compensation for their participation. The reduced sample of 40 participants (see Data treatment and analyses for details on data exclusion) provides a power of 80% to detect an effect of $d_z = 0.45$ ($\alpha = 5\%$, two-tailed).

Apparatus and Stimuli

The apparatus was the same as in Experiment 1, but the task and stimulus material differed. In a prime task, participants categorized target numbers (i.e., 1, 2, 3, 4, 6, 7, 8 and 9) as odd or even through key presses with their index fingers (keys F and J). This procedure replaces the more common free-choice setup of previous studies (e.g., Dutzi & Hommel, 2009; Schwarz et al., 2018) in order to induce errors in the prime segment—which can only be detected for forced-choice setups. Note that short-term binding
between responses and effects arises also for such forced-choice actions (Herwig & Waszak, 2012; Janczyk et al., 2012). The assignment of the two number categories to the keys was counterbalanced across participants. The target number appeared centrally in white font against a black background. Above, below, left and right of the target appeared an irrelevant letter in white font that was drawn from a pool of eight letters (i.e., O, W, X, U, Z, Y, H, A). The same letter was displayed at all four positions (horizontal displacement: 115 px, vertical displacement: 108 px, relative to the center of the target number). Again, we included these letters to increase commission errors through perceptual noise. Prime responses randomly produced a 400 Hz or 800 Hz sinusoidal effect sound for 300 ms. The probe task featured an exclamation mark in white font on the screen and a 400 Hz, 600 Hz or 800 Hz sound for 300 ms. Participants were to choose between the left and the right key when they heard the 400 Hz or 800 Hz sound but they were to refrain from responding for the 600 Hz sound.

Procedure

Participants first got to know about the mapping rule of the prime task and that their responses would randomly produce one of two sounds. They then listened to both sounds once and learned that they would hear another sound afterward and that this sound could be one of the two sounds they just heard. In this case, they were to choose one of the keys spontaneously without using a strategy (German original: “aus dem Bauch heraus”). If they heard a third sound, they would have to make no keypress. They listened to that third sound after these instructions. Finally, participants read a summary of the trial procedure and were urged to deliver their second response in the trial even if they failed to deliver a correct first response. Participants could either proceed to the practice block or could decide to go over the instructions again.

Each trial began with a white fixation cross in the center of a black screen for 500 ms (see Figure 3). Afterward, the prime number and irrelevant letters stayed on screen until participants responded or 600 ms had passed. The effect sound was played back as soon as participants delivered a prime response whereas there was no sound if they did not respond to the prime. In the practice block, participants received feedback on each prime response for 1000 ms: “Correct!” (German original: “Richtig!”) in green font color for correct responses, “Too slow!” (German original: “Zu langsam!”) in red font color for omission errors, and “Wrong!” (German original: “Falsch!”) in red font for commission or random errors. In all following blocks, participants only received immediate feedback for omission errors. They received feedback about the percentage of correct responses and their mean correct response time in the prime at the end of each block. Again, the fixation cross appeared on screen for 725 ms, followed by a white exclamation mark in the center of the screen. The probe sound accompanied the exclamation mark for 300 ms. If the probe sound was a 400 Hz or 800 Hz sound, participants were to deliver a free-choice response with the instructed keys. The 600 Hz sound marked a catch trial and participants had to refrain from responding. Participants received error feedback in red font for 1000 ms if they responded in catch trials (“Do not respond to this sound!”; German original: “Bei diesem Ton nicht reagieren!”) or used any other than the instructed keys in the free-choice trials (“Wrong!”; German original: “Falsch!”). After each block, participants received a reminder to respond to the second sound as spontaneously as possible.

Figure 3
Exemplary Trial Procedure and Experimental Conditions of Experiment 2

Note. In the prime task, participants had to respond to the central target letter within 600 ms and they had to ignore the surrounding letters. An effect sound (400 Hz vs. 800 Hz) played for 300 ms after their response. After another fixation, a probe sound (400 Hz vs. 800 Hz) played for 300 ms with an exclamation mark on screen until participants responded. In catch trials, the probe task featured a third sound (600 Hz) upon which participants had to refrain from responding in a time window of 3000 ms from the onset of the sound.
Participants went through one practice block and 18 experimental blocks that each featured 32 prime-probe sequences. Prime numbers and letters were shuffled independently after each eighth trial without repetitions of either stimulus in successive trials. As such, each prime number and each letter appeared four times in a block. Half of the trials of a block presented the 400 Hz sound as prime effect whereas the 800 Hz sound appeared as effect in the other half of the trials, with both sounds appearing in a random sequence. The probe sound was chosen pseudorandomly to allow for sufficient cell observations of sound repetitions and changes even for a small number of errors. If participants made a commission error in the prime, the probe sound was chosen from an array that held a random sequence of two elements coding for sound repetitions from prime to probe, two elements coding for sound changes from prime to probe and one catch trial where participants had to refrain from responding in the probe. In catch trials, the probe sound was always 600 Hz. For sound repetitions and changes, the probe sound depended on the prime sound. For

**Figure 4**

*Theoretical Predictions and Main Results of Experiment 2*

(A) Potential mechanisms for binding the representation of an erroneous (R_{err}) or a correct response (R_{cor}) to the effect of the erroneous response (E_{err}). Dashed outlines represent binding of an event file. *Predictions* (shapes refer to sequential conditions as shown in Panel B): For action slips (bright, orange), the *success-based* account again assumes no binding, predicting the same proportion of response repetitions from a forced-choice prime response to a free-choice probe response for action sequences where the effect sound of the prime response and the target sound of the probe are the same (sound repetitions; triangle) and sequences where the sound changes from prime to probe (square). The *co-activation* account assumes binding between the erroneous response and the erroneous effect, predicting more response repetitions if the sound repeats. The *goal-based* account instead predicts a higher frequency of response repetitions if the sound changes because of binding between the correct response and the erroneous effect. For correct action episodes (grey/black), all accounts predict increased response repetitions for sound repetitions. (B) The results support binding through co-activation for action slips. Error bars represent the standard errors of paired differences (SEPD), computed separately for sound repetitions and changes. (C) Individual binding effects (dots) in the proportion of response repetitions with their estimated density distribution for preceding correct and erroneous responses. Lines connect the two binding effects of correct and erroneous episodes for each participant. Bimodality coefficients (BCs) point to unimodal distributions for both types of responses (BC\_crit = .555). See the online article for the color version of this figure.

**Note.**  
(A) Potential mechanisms for binding the representation of an erroneous (R_{err}) or a correct response (R_{cor}) to the effect of the erroneous response (E_{err}). Dashed outlines represent binding of an event file. *Predictions* (shapes refer to sequential conditions as shown in Panel B): For action slips (bright, orange), the *success-based* account again assumes no binding, predicting the same proportion of response repetitions from a forced-choice prime response to a free-choice probe response for action sequences where the effect sound of the prime response and the target sound of the probe are the same (sound repetitions; triangle) and sequences where the sound changes from prime to probe (square). The *co-activation* account assumes binding between the erroneous response and the erroneous effect, predicting more response repetitions if the sound repeats. The *goal-based* account instead predicts a higher frequency of response repetitions if the sound changes because of binding between the correct response and the erroneous effect. For correct action episodes (grey/black), all accounts predict increased response repetitions for sound repetitions. (B) The results support binding through co-activation for action slips. Error bars represent the standard errors of paired differences (SEPD), computed separately for sound repetitions and changes. (C) Individual binding effects (dots) in the proportion of response repetitions with their estimated density distribution for preceding correct and erroneous responses. Lines connect the two binding effects of correct and erroneous episodes for each participant. Bimodality coefficients (BCs) point to unimodal distributions for both types of responses (BC\_crit = .555). See the online article for the color version of this figure.
example, sound repetitions would call for a 400 Hz probe and sound changes for an 800 Hz probe if the prime effect was 400 Hz. If all elements had been drawn from the array, the array was randomized (i.e., after each 5th commission error). A second array with the same elements was used for all other kinds of prime responses (i.e., correct responses, omissions, and random errors).

**Data Treatment and Analysis**

We decided to exclude two participants although the reasons of exclusion had not been anticipated in the preregistration. One participant responded with only one hand instead of their two index fingers, which the experimenter only noticed at the end of the experiment. The second participants responded multiple times before the onset of the probe in most trials (53.8%).

We eliminated the practice block and all catch trials. Participants correctly refrained from responding to the catch probe in about 72% of the catch trials. We then filtered trials with miscellaneous errors (3.4%) or an omission error in the prime (7.4%). Finally, we filtered trials with probe responses with any other than the instructed keys (0.1%). Six participants had to be excluded after this data selection because they provided less than ten observations in at least one of the experimental cells. The remaining 40 participants delivered on average 179 trials ($SD = 23$) of sound repetitions and 179 trials ($SD = 21$) of sound changes after a correct prime response. After a commission error in the prime, they delivered on average 27 trials ($SD = 14$) of sound repetitions and 27 trials ($SD = 14$) of sound changes.

We analyzed the relative frequency of response repetitions from prime to probe in a $2 \times 2$ ANOVA with the within-subject factors prime accuracy (correct vs. commission error) and sound sequence (repetition vs. change). With the same selection of trials and participants, we analyzed response times in the probe as a function of the accuracy of the prime response in a two-tailed paired-samples t-test. We did not include the factor sound sequence because response times are typically noisy in such designs (Dutzi & Hommel, 2009; Moeller, Pfister, et al., 2016).

**Results**

Figure 4 shows the predictions of the three theoretical accounts alongside the main results of Experiment 2 (see Table A2 in the Appendix for a full overview of the descriptive data for each dependent variable and experimental cell).

**Frequency of Response Repetitions**

Correct and erroneous prime responses were repeated more frequently in the probe if the sound repeated rather than changed (Figure 4B), $F(1, 39) = 18.47, p < .001$, $\eta^2_g = .32$. Although binding effects did not differ between correct ($M = 10\%$, $SD = 11\%$) and erroneous action episodes ($M = 5\%$, $SD = 17\%$), $F(1, 39) = 2.43, p = .127$, $\eta^2_p = .06$, response repetitions occurred more frequently after correct than after erroneous responses, $F(1, 39) = 32.45, p < .001$, $\eta^2_p = .45$.

**Response Times and Distributional Analyses**

Response time analyses pointed to a descriptive trend of post-error slowing in the probe ($M = 30$ ms, $SD = 97$), $t(39) = 1.98, p = .055$, $d = 0.31$, 95% CI $[-1$ ms, 61 ms]. Follow-up correlational analyses did not show a significant relation between binding effects of correct and erroneous episodes across participants (Figure 4C), $r = .26, t(42) = 1.69, p = .099$. The distribution of both binding effects was unimodal (bimodality coefficients $\leq .511$).

**Discussion**

The results of Experiment 2 reveal that the representation of response-triggered effects in the environment is bound to the action that produced them, irrespective of whether this action unfolded as planned or whether it was committed in error. From prime to probe, sound repetitions increased response repetitions compared to sound changes. As in Experiment 1, the distributional analyses again pointed to a single binding mechanism in erroneous action episodes. The similarity of binding effects for executed erroneous and correct responses to their effects supports our preceding argument that binding of the correct response to the target is less rich for erroneous than correct action episodes because execution-specific features can only be predicted in this case. Somewhat surprising in this regard is that the correlation between response-effect bindings of correct and erroneous action episodes was not significant. Larger sample sizes might uncover any small relation between the effects.

The results indicate binding through co-activation even though participants detected their errors—that is, responses were not just misperceived as being correct: Errors came with systematic aftereffects in that participants repeated erroneous response less often than correct responses, suggesting that errors were indeed registered and actively avoided (Aarts et al., 2012). We anticipated that response times would be noisy in such designs (Dutzi & Hommel, 2009; Moeller, Pfister, et al., 2016) and are thus not concerned about the descriptive trend, but nonsignificant post-error slowing effect.

**General Discussion**

Together, the two experiments demonstrate that the human cognitive system binds action and perception with striking precision. When actions go awry, agents correct traces of error commission on the fly through binding of the intended correct response to the previous target instead of the actually executed erroneous response (for binding of irrelevant stimulus features to intended correct responses, see Foerster et al., in press). At the same time, responses can still be bound to the effects that they produce, irrespective of the appropriateness of the response. These two findings characterize binding as surprisingly adaptive, because binding creates compounds that are tuned to the agent’s environment by reinforcing the previous action goal while at the same time integrating contiguous responses and effects. The expansion of binding from successful to unsuccessful action episodes therefore suggests that binding sensorimotor features supports error-based learning on a

---

4 We also approached binding of responses and their effects via forced-choice responses in the probe in two additional experiments. The first experiment had high data exclusion rates and both experiments did not replicate the traditional binding effect for correct responses. This precludes any inferences about binding for action slips. Preregistrations of both experiments and a thorough report can be accessed in the Open Science Framework project that also provides the data of Experiment 1 and 2 (osf.io/3at7x).
basic level of perceptual and action representations (Metcalfe, 2017).

The current study investigated bindings of targets and responses (Experiment 1) as well as binding of responses and their effects (Experiment 2). To do so, we employed distinct designs that have been established as tried-and-tested methods in the corresponding scientific communities. Two recent studies further suggest that these designs tap into the same mechanisms while suggesting that stimulus-response and response-effect relations coexist independently (Moeller, Frings, et al., 2016, 2019). This might be taken to suggest that both types of binding are indeed compiled for each action episode. Alternatively, the cognitive system might employ only one type of binding mechanism depending on situational constraints and potential context factors. Future work should therefore build on these insights by developing similar integrative designs also in the context of binding for action slips, ideally allowing for measuring goal-based binding of stimulus and response as well as binding through co-activation in case of action effects within the same trial.

The potential operation of both binding mechanisms within the same action episode requires triggers that indicate the need for one or the other mechanism. The prediction error inherent in action slips might serve as a signal to structure the stream of events into relevant units (Botvinick et al., 2001; Rouhani et al., 2020; Verguts & Notebaert, 2009; Zacks et al., 2007). This signal should be available promptly as indicated by the electrophysiological signature of errors, with distinct event-related potentials that map to common frontal midline theta activity (e.g., Cavanagh & Frank, 2014; Falkenstein et al., 2000; Hoffmann & Beste, 2015). Interestingly, recent studies demonstrated that binding and retrieval effects emerge in theta activity as well, for example modulating the efficiency of the organization of the underlying networks (Takacs et al., 2020). Alternatively, goal-based binding might only incorporate stimuli that appear to be related to the planning of goal-directed behavior whereas co-activation would consider stimuli that only appear because of actions. This perspective considers the response itself, be it correct or incorrect, as a temporal or causal source of event segmentation that allows for different binding mechanisms.

In any case, the nature of feature binding in the cognitive system thus appears both simple and powerful: By being tuned to perception and action contingencies, it promotes efficient and successful behavior in the future.

Context

The presented experiments emerged from a project that is dedicated to testing the three proposed accounts of binding for action slips across error types and for a variety of features of an action episode. The project is part of a Research Unit of the German Research Foundation (“Binding and Retrieval in Action Control”, FOR 2790), aiming at developing a broad theoretical framework that puts binding and retrieval at the heart of theorizing on action control and related fields (Frings et al., 2020). Action slips have so far proven to be an excellent testbed for this endeavor and the studies further inspired us to see instant error processing as a surprisingly dynamic and adaptive process.

References


Appendix

Descriptive Statistics

Table A1

Descriptive Data of Experiment 1

<table>
<thead>
<tr>
<th>Preceding accuracy</th>
<th>Condition sequence</th>
<th>RT in ms</th>
<th>Commissions in %</th>
<th>Omissions in %</th>
<th>RV (_{RT}) in ms</th>
<th>(\Delta RD) in ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>Target repetition</td>
<td>correct response repetition</td>
<td>370 (44)</td>
<td>2.6 (4.5)</td>
<td>1.3 (1.4)</td>
<td>18 (6)</td>
</tr>
<tr>
<td></td>
<td>Target change</td>
<td>correct response repetition</td>
<td>405 (57)</td>
<td>11.4 (6.3)</td>
<td>5.6 (4.1)</td>
<td>16 (12)</td>
</tr>
<tr>
<td>Error</td>
<td>Target repetition</td>
<td>correct response repetition</td>
<td>453 (46)</td>
<td>25.7 (14)</td>
<td>7.0 (5)</td>
<td>11 (18)</td>
</tr>
<tr>
<td></td>
<td>Target change</td>
<td>correct response change</td>
<td>419 (64)</td>
<td>11.7 (12.4)</td>
<td>5.8 (6.4)</td>
<td>16 (15)</td>
</tr>
<tr>
<td></td>
<td>Target change</td>
<td>correct response change</td>
<td>430 (63)</td>
<td>15.3 (11.4)</td>
<td>8.3 (5.6)</td>
<td>17 (19)</td>
</tr>
</tbody>
</table>

Note. Mean response times (RTs) in ms, commission and omission errors in %, relative variance of response times (RV\(_{RT}\)) in ms, and absolute differences in response duration (\(\Delta RD\)) between succeeding trials in ms with respective standard deviations (SD) in brackets for each combination of preceding accuracy and condition sequence.

Table A2

Descriptive Data of Experiment 2

<table>
<thead>
<tr>
<th>Prime accuracy</th>
<th>Sound sequence</th>
<th>Response repetitions in %</th>
<th>Response times in ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>Repetition</td>
<td>71 (26)</td>
<td>521 (132)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>61 (27)</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>Repetition</td>
<td>52 (21)</td>
<td>569 (200)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>47 (26)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Mean response repetitions in % and response times in ms with respective standard deviations (SD) in brackets for each combination of prime accuracy and sound sequence.

Figure A1

Error Results of Experiment 1

Note. Mean percentage (A) commission and (B) omission errors for all sequential conditions. Error bars represent within-subject standard errors for the full factorial design (SE\(_{Within}\)). See the online article for the color version of this figure.

(Appendix continues)
Figure A2

*Bimodality Coefficients of Experiment 1*

*Note.* Individual (dots) and mean (triangles) bimodality coefficients (BCs) of response times for trial sequences with repetitions of the target and the correct response. BCs of correct (grey/black) and erroneous (bright orange) action episodes point to unimodal distributions (BCcrit = .555). Error bars represent the standard errors of means (SEM). See the online article for the color version of this figure.