Cogn Affect Behav Neurosci (2012) 12:323–336
DOI 10.3758/s13415-012-0087-9

Feedback-related brain activity predicts learning from feedback in multiple-choice testing

Benjamin Ernst · Marco Steinhauser

Published online: 13 March 2012

Abstract Different event-related potentials (ERPs) have been shown to correlate with learning from feedback in decision-making tasks and with learning in explicit memory tasks. In the present study, we investigated which ERPs predict learning from corrective feedback in a multiple-choice test, which combines elements from both paradigms. Participants worked through sets of multiple-choice items of a Swahili–German vocabulary task. Whereas the initial presentation of an item required the participants to guess the answer, corrective feedback could be used to learn the correct response. Initial analyses revealed that corrective feedback elicited components related to reinforcement learning (FRN), as well as to explicit memory processing (P300) and attention (early frontal positivity). However, only the P300 and early frontal positivity were positively correlated with successful learning from corrective feedback, whereas the FRN was even larger when learning failed. These results suggest that learning from corrective feedback crucially relies on explicit memory processing and attentional orienting to corrective feedback, rather than on reinforcement learning.

Keywords Feedback · Dm effect · P300 · Multiple-choice testing · Event-related potentials

Introduction

Multiple-choice testing is a frequently used examination method in school and university education. It requires that students respond to a probe (e.g., a question like "What is the capital of France?") by choosing among a set of alternatives including a target and several distractors (e.g., "Berlin," "Paris," "London," "Rome"). This method has a number of advantages over other testing procedures: The results can be graded objectively, easily, and quickly. Moreover, not only are multiple-choice tests helpful for probing knowledge, they can also be used as training material, because testing itself improves the retention of the material tested—the well-known testing effect (Pyc & Rawson, 2010; Roediger & Karpicke, 2006; Spitzer, 1939). To enable learning from multiple-choice tests, it is crucial that incorrect responses are followed by corrective feedback containing information about the correct response. In the present study, we used event-related potentials (ERPs) to investigate which processes are involved when participants learn from corrective feedback in a multiple-choice test.

Whereas the beneficial effect of feedback in multiple-choice testing has previously been demonstrated (e.g., Butler & Roediger, 2008), it is still unclear which processes are crucially involved when participants learn from corrective feedback in a multiple-choice test. Traditionally, the role of feedback in learning has been discussed from two perspectives. In literature on memory and verbal instruction, it has been argued that feedback primarily serves to correct errors by providing information about the correct response (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kulhavy, 1977). From this perspective, feedback processing resembles the processing of the learning material itself and, thus, mainly involves memory encoding and elaboration. In contrast, the literature on decision making has emphasized the role of feedback as a reinforcer (Frank & Claus, 2006; Holroyd & Coles, 2002). From this perspective, feedback triggers the strengthening or weakening of stimulus–response associations underlying the previous decision.

B. Ernst (✉) · M. Steinhauser
Department of Psychology, University of Konstanz,
Fach D29,
78457 Konstanz, Germany
e-mail: Benjamin.Ernst@uni-konstanz.de

Crucially, both perspectives can be applied to multiple-choice testing. A multiple-choice test can be viewed as a decision-making paradigm in which a decision is made by retrieving information from memory. Corrective feedback conveys two types of information, potentially serving two functions. On the one hand, it conveys information about the correct response, which can be encoded to replace incorrect information on the basis of declarative memory processing. On the other hand, it conveys information about the correctness of the response, which can be used to strengthen or weaken a previous decision on the basis of reinforcement learning. Indeed, it is conceivable that both processes are involved in multiple-choice testing. Recent models of categorization and decision making propose that learning from feedback involves both implicit and explicit memory processes (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Frank & Claus, 2006). In the present study, we investigated the contribution of these different processes to learning from feedback by considering ERP components that are known to be related to memory encoding and feedback processing. In the next sections, we provide a brief overview of relevant ERP components from both fields.

ERPs related to memory processing

Initial studies on ERP correlates of memory encoding focused on the P300. The *stimulus-locked P300* is a positivity emerging around 300 ms after stimulus onset and consists of an earlier frontal subcomponent (P3a) and a later parietal subcomponent (P3b). Whereas the frontal P3a has been assumed to reflect an attentional orienting response to a novel or unexpected stimulus (Simons, Graham, Miles, & Chen, 2001), the parietal P3b has been thought to be related to memory (Donchin, 1981; Polich, 2007; Polich & Kok, 1995). To investigate which ERPs are involved in memory encoding, several studies analyzed ERPs during learning of, for example, word lists or paired associates, as a function of whether an item was subsequently retrieved or not. Paller, Kutas, and Mayes (1987) proposed the term "dm" to refer to these ERP differences that reflect later memory performance.

Dm effects have been found in a large number of ERP components (Fabiani, Karis, & Donchin, 1986, 1990; Friedman, Nessler, & Johnson, 2007; Karis, Fabiani, & Donchin, 1984; Kim, Vallesi, Picton, & Tulving, 2009; Mangels, Picton, & Craik, 2001; Paller et al., 1987; Sommer, Heinz, Leuthold, Matt, & Schweinberger, 1995; Weyerts, Tendolkar, Smid, & Heinze, 1997). Two of them were obtained rather consistently across studies. First, there was an increased parietal positivity for successfully encoded material in the time range of the P300 (Fabiani et al., 1986, 1990; Karis et al., 1984; Kim et al., 2009; Mangels et al., 2001). Second, some studies found an increased *late frontal positivity* for successfully encoded material that typically started 600 ms or later after stimulus onset and lasted for several hundreds of milliseconds (Fabiani et al., 1990; Kim et al., 2009; Mangels et al., 2001; Weyerts et al., 1997). There is some evidence that the P300 effect is related more to rote learning based on physical characteristics of the stimuli, whereas the late frontal positivity reflects more elaborate processing (Fabiani et al., 1990; Kim et al., 2009).

ERPs related to feedback processing

Feedback processing has mainly been investigated using simple decision tasks in which feedback indicated only the correctness of the response. The most frequently examined component in these studies is the *feedback-related negativity* (FRN), a negative deflection with a maximum at about 250 ms following feedback onset at fronto-central electrode sites. The FRN amplitude is larger for negative feedback than for positive feedback, indicating that the FRN is sensitive to feedback valence (Gehring & Willoughby, 2002; Holroyd & Coles, 2002; Luu, Tucker, Derryberry, Reed, & Poulsen, 2003; Miltner, Braun, & Coles, 1997; Yeung, Holroyd, & Cohen, 2005). A current theory of the FRN assumes that it reflects a reinforcement signal that triggers learning from negative feedback (Holroyd & Coles, 2002), which explains why FRN amplitude (Bellebaum & Daum, 2008; Cohen & Ranganath, 2007; Philiastides, Biele, Vavatzanidis, Kazzer, & Heekeren, 2010; van der Helden, Boksem, & Blom, 2010) or oscillatory activity related to the FRN (Cavanagh, Frank, Klein, & Allen, 2010) predicts the strength of behavioral adjustments in simple decision tasks.

Feedback in decision-making tasks typically has elicited a second component that overlaps in time with the FRN. This *feedback-locked P300* refers to a positivity over parietal electrode sites that reaches a maximum about 300 ms after feedback onset. It is sometimes reported to be less sensitive to feedback valence (Yeung & Sanfey, 2004; but see Bellebaum & Daum, 2008; Bellebaum, Polezzi, & Daum, 2010; Hajcak, Moser, Holroyd, & Simons, 2007; van der Helden et al., 2010) but—at least in simple decision tasks—correlates with reward magnitude (Yeung & Sanfey, 2004) and reward probability (Hajcak, Holroyd, Moser, & Simons, 2005; Hajcak, Moser, Holroyd, & Simons, 2006; San Martin, Manes, Hurtado, Isla, & Ibanez, 2010) and, therefore, has been assumed to reflect the update of outcome expectancies in these tasks (e.g., Yeung & Sanfey, 2004). Recently, Chase, Swainson, Durham, Benham, and Cools (2011) showed that explicit rule-based adjustment in a probabilistic reversal learning task is predicted by the P300, whereas the FRN is linked only to reinforcement learning. This suggests that learning in decision-making tasks is based on explicit and implicit memory processes (Frank & Claus, 2006), with the P300 being more related to the former and the FRN being more related to the latter. This has received further support from studies finding no relation between FRN amplitude and learning from feedback in tasks in which explicit learning is involved (Butterfield & Mangels, 2003; Mangels, Butterfield, Lamb, Good, & Dweck, 2006, Mangels, Good, Whiteman, Maniscalco, & Dweck, 2011).

Whereas the FRN and feedback-locked P300 have typically been investigated for feedback indicating only the correctness of the response, fewer studies have considered corrective feedback—that is, feedback containing information about the correct response. For instance, Butterfield and Mangels (2003) examined learning from feedback in a semantic retrieval task, in which participants answered simple questions followed by corrective feedback. They found an *early frontal positivity* following negative feedback peaking at about 350 ms following feedback onset over fronto-central electrode sites when participants successfully learned from this feedback. Because this component varied with the expectedness of negative feedback, Butterfield and Mangels assumed that it is functionally related to a P3a. When observed following corrective feedback, this frontal positivity could reflect attentional orienting to information about the correct response that occurs when negative feedback is detected (Butterfield & Mangels, 2003).

Taken together, our brief overview has revealed that feedback processing and memory encoding are reflected by partially overlapping sets of ERP components. Encoding an item in explicit memory is related to the amplitude of the P300 and—if more elaborate processing takes place—to the amplitude of a late frontal positivity. Feedback processing has been shown to involve an FRN that reflects reinforcement learning, a P300 that reflects feedback processing in working memory, and, in the case of corrective feedback, an early frontal positivity that could be related to attentional orienting. The question emerges as to which of these components predict successful learning from corrective feedback in multiple-choice testing, a paradigm that combines characteristics of a decision-making task with characteristics of a memory task.

We hypothesize that successful learning should be related to the FRN if learning relies on reinforcement, which is possible because a multiple-choice test is a decision-making task in which an association between a stimulus and a response has to be learned. In contrast, successful learning should be related to a P300 or a late frontal positivity if learning relies on explicit memory encoding mediated by working memory processes, which is possible because multiple-choice tests also require retrieving information from explicit memory. Finally, on the basis of the results of Butterfield and Mangels (2003), we hypothesize that successful learning should be related to an early frontal positivity related to attentional orienting toward information about the correct response.

The present study

The goal of the present study was to investigate which ERP components predict successful learning from corrective feedback in multiple-choice testing. To achieve this, we analyzed feedback-related ERPs in a Swahili–German vocabulary task in which a Swahili probe word was presented and participants had to choose the correct German translation from a set of four alternatives. Feedback was provided indicating the correctness of the response together with the correct translation. Each set of items was presented in two blocks (see Fig. 1) in which participants had to learn the correct response by responding and then processing corrective feedback. In the first learning block—the guess-and-learn block—participants were naive about the correct translations, and performance was exclusively due to guessing. In the second learning block—the test-and-learn block—the same probe items were presented again, and performance should reflect the efficiency by which feedback was employed to learn the correct translation in the guess-and-learn block.

Data from these two blocks were used for investigating which components of feedback processing were predictive of successful learning from feedback. To achieve this, feedback-related ERP data in the guess-and-learn block were analyzed as a function of performance in the test-and-learn block. We distinguished between *E–C items,* for which the initial response in the guess-and-learn block was an error (E) but which were answered correctly in the test-and-learn block (C), indicating that the participants learned from feedback, and *E–E items,* for which the response in the guess-and-learn block was incorrect (E) and which were

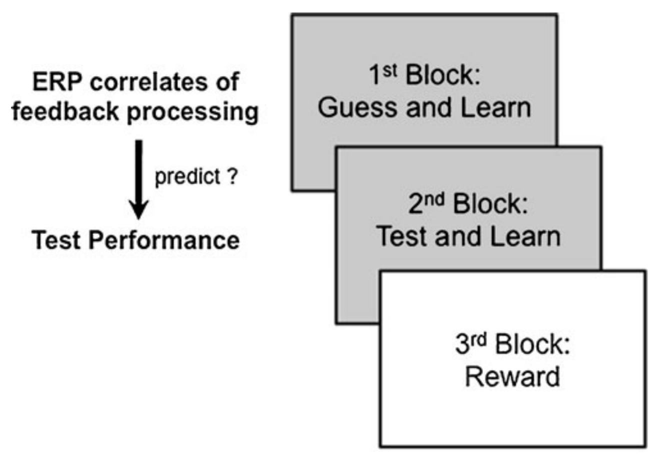

Fig. 1 Block types in the experiment. Each item was presented in a guess-and-learn block, in which participants guessed and then learned from feedback, in a test-and-learn block, in which the same items were presented and participants again learned from feedback, and in a reward block, in which participants were finally tested and received rewards according to their performance

also answered incorrectly in the test-and-learn block (E), indicating that participants failed to learn from feedback. The two learning blocks were followed by a reward block, in which performance was associated with monetary wins and losses for correct and incorrect responding.[1]

Method

Participants

Thirty-six participants (23 females) between 18 and 29 years of age (mean, 22.5) with normal or corrected-to-normal vision participated in the study. All participants were recruited at the University of Konstanz. Whereas half of the participants received a base fee of 20 € and a small performance-dependent bonus (mean, 0.89 €; range, 0.19 €–1.18 €), the other half received a smaller base fee of 15 €, but a larger and more variable performance-dependent bonus (mean, 7.31 €; range, 0.30 €—11.90 €).[2] The study was conducted in accordance with institutional guidelines, and informed consent was acquired from all participants.

Stimulus materials

The words were 240 German nouns and their proper translations in Swahili and an additional 240 German nouns that were used only as distractors and, therefore, had no Swahili equivalent. German nouns were disyllabic and abstract and had a mean word frequency of 199 per million (*SD* = 222 per million) (based on CELEX Lexical Database of the Dutch Centre for Lexical Information; see Burnage, 1990). The Swahili words were disyllabic or trisyllabic. Examples are ABEND/JIONI (English, *evening*), GEDULD/SUBIRA (English, *patience*), or THEMA/INSHA (English, *topic*). For each participant, the words were randomly subdivided into eight target lists, T_1–T_8, and eight distractor lists, D_1–D_8.

At the beginning of each block, 30 new items were generated that consisted of a Swahili probe, the corresponding German target, and three German distractors. For each target from list T_i, three distractors were chosen by randomly sampling one or two distractors from distractor list D_i (with replacement) and one or two distractors from target list T_i (with replacement). This procedure ensured that the items could not be solved by merely recognizing words from the target list. It was rather necessary to identify the target associated with the probe.

All words were taken from Arial font and had a height of 0.4° of visual angle and a width between 0.8° and 1.6° of visual angle at a viewing distance of about 70 cm. Stimuli were presented in white color on a black background and consisted of a centrally presented Swahili probe and four German words (a target and three distractors). The German words were located at the corners of an invisible square (4.7° × 4.7°) around the probe (see Fig. 2).

Design and procedure

Items were presented in different types of blocks. In *guess-and-learn* blocks and *test-and-learn* blocks, the stimulus was presented, and participants were instructed to indicate the location of the target word by pressing one of four keys on a German standard keyboard (A with left middle finger, Y with left index finger, K with right middle finger, M with right index finger) within 5 s. After 4 s, the stimulus turned gray to inform participants that only 1 s was left. Following a response (or after 5 s), the German words disappeared, and the Swahili probe remained on the screen for another 1,500 ms. This was done to trigger refixation to the central word before feedback presentation. The feedback consisted of the centrally presented target word. The word color signaled the accuracy of the response, with red indicating an incorrect response and green indicating a correct response. The feedback remained on the screen for 5 s, followed by 2 s with a blank screen, and participants were instructed to use this time for memorizing the correct translation.

In *reward blocks*, stimulus presentation and responding was the same as in the guess-and-learn/test-and-learn blocks, with one exception. Feedback was presented only for 1,500 ms, because no further learning was necessary. Participants were paid according to their performance in reward blocks only. In half of the participants, correct responses were associated with a bonus of 1 ct, incorrect responses were associated with a malus of 1 ct, and misses were associated with a malus of 3 ct. In the other half of the participants, the corresponding values were +10 ct, −10 ct, and −30 ct.

Participants worked through four item lists, and the order of items was newly randomized before each block. Each item list was learned in a guess-and-learn block and a test-and-learn block and was then tested in a reward block. This resulted in 12 blocks with 30 trials each, which were presented in one session lasting for approximately 1 h. There was a break of 10 s between blocks. After guess-and-learn, test-and-learn, and reward blocks, a summary of correct,

[1] Our main analysis required that the task be sufficiently difficult to obtain a very high number of incorrect responses in the second block. Therefore, we decided to include a final test block to provide participants with the opportunity to achieve a satisfying learning result.

[2] Robust effects of this incentive manipulation were obtained neither for the overall learning performance nor for the analysis of ERP data. There was only a trend toward an increased early frontal positivity in the condition with lower incentives. Therefore, this manipulation was not further discussed.

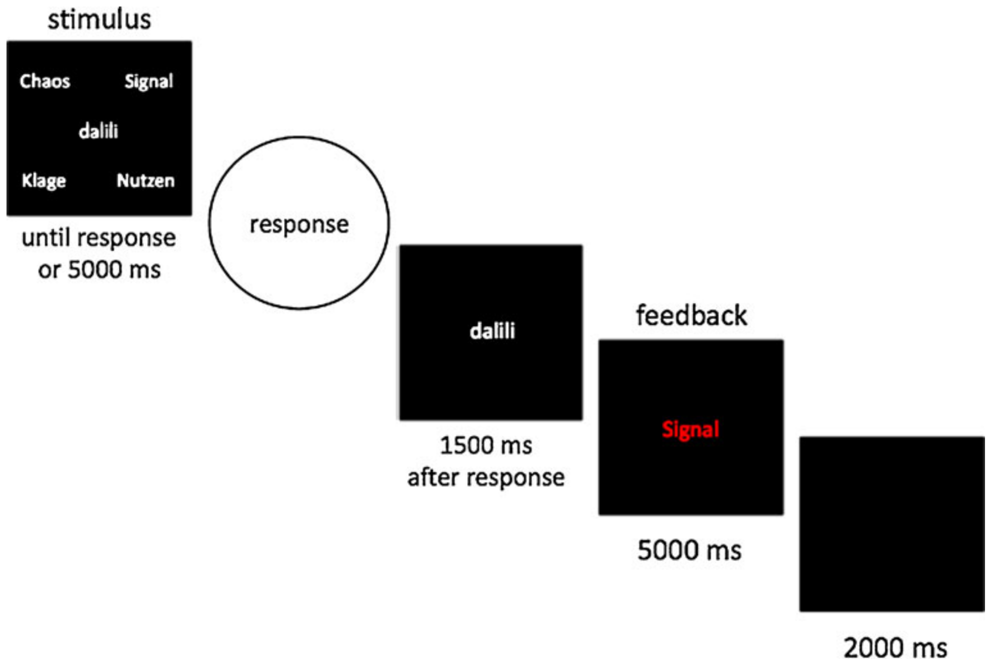

Fig. 2 Sequence of events in a typical trial of a guess-and-learn or test-and-learn block. The stimulus consisted of a Swahili probe and four German translations. After participants pressed a response key, the probe was presented in isolation, followed by a feedback stimulus indicating the correct translation as well as the correctness of the participants' response (color green = correct; color red = incorrect)

incorrect, and missed responses for this block was provided for 10 s. After each reward block, information about the total reward was provided, and participants had a short break. In addition to this session, participants were also tested in a second session, in which different word pairs had to be learned but in which participants were instructed not to respond during the first 2 blocks. The data from this session cannot be used to investigate feedback processing, because no feedback about a response was provided, and, therefore, are not reported in the present article. The order of sessions was counterbalanced across participants. They took place on two different days, separated by 1 week, each at approximately the same time of day.

Electrophysiological recordings

Throughout the experiment, participants were seated comfortably in a dimly lit, electrically shielded room. The electroencephalogram (EEG) was recorded using a BIOSEMI Active-Two system (BioSemi, Amsterdam, The Netherlands) with 64 Ag-AgCl electrodes from channels Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C1, C3, C5, T7, TP7, CP5, CP3, CP1, P1, P3, P5, P7, P9, PO7, PO3, O1, Iz, Oz, POz, Pz, CPz, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz, Cz, C2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, and O2, as well as the left and right mastoid. The CMS (common mode sense) and DRL (driven right leg) electrodes were used as reference and ground electrodes. Vertical and horizontal electrooculogram (EOG) was recorded from electrodes above and below the right eye and on the outer canthi of both eyes. All electrodes were offline rereferenced to averaged mastoids. EEG and EOG were continuously recorded at a sampling rate of 512 Hz.

Data analysis

The first word list was regarded as practice and was excluded from analyses. Moreover, trials were excluded on which participants did not respond or on which responses were too slow (mean: 3.83 trials per participants).

EEG data were analyzed using EEGLAB v6.01 (Delorme & Makeig, 2004) and custom routines written in MATLAB 7.0.4 (The Mathworks, Natick, MA). The data were bandpass filtered excluding activity below 1 Hz and above 30 Hz. Epochs ranging from 100 ms before to 500 ms after feedback onset were extracted. Large artifacts were identified by computing the joint probability of each epoch and by excluding epochs that deviated more than five standard deviations from the distribution mean and epochs in which activity exceeded a threshold of ± 150 µV. Further artifacts were removed by applying independent component analysis and eliminating components that were identified as reflecting ocular or muscle activity (cf. Delorme, Sejnowski, & Makeig, 2007). Baseline activity was removed by subtracting the average voltage from an interval between 100 and 0 ms before feedback onset. Finally, epochs were averaged separately for each condition of interest.

Two methods were used to quantify feedback-locked ERPs. First, we extracted mean amplitudes from four equally long time windows (I, 265–308 ms; II, 312–355 ms; III, 359–402 ms; IV, 406–449 ms). The time windows were chosen on the basis of visual inspection of the waveforms for positive and negative feedback. First, components of interest were identified (feedback-locked P300, FRN, frontal positivity), and then positions and width of the time windows were selected so that each component of interest was centered within and covered by a set of time windows (feedback-locked P300 in windows I–IV, FRN in windows I and II, frontal positivity in window III). Mean amplitudes were subjected to repeated measurement ANOVAs with within-subjects variables of time window, electrode, and condition of interest. Second, we additionally applied a peak-to-peak measure to quantify the FRN (e.g., Yeung & Sanfey, 2004). Peak-to-peak voltages of the FRN were quantified by first filtering the data with a 10-Hz low-pass filter and then calculating the difference between the most negative peak between 200 and 400 ms after feedback onset and the immediately preceding positive peak. This was done separately for each participant and each condition of interest. FRN amplitudes were subjected to repeated measurement ANOVAs with the within-subjects variables of electrode and condition of interest. Although we provide data for larger sets of electrodes in the figures, we restricted statistical analyses to electrodes FCz (or Fz) and Pz, which are representative for anterior and posterior effects in our data. To compensate for violations of sphericity, Huynh–Feldt corrections were applied whenever appropriate, and corrected p-values (but uncorrected degrees of freedom) are reported.

Results

Behavioral data

Behavioral data were analyzed to investigate whether participants were able to learn from feedback in our paradigm. We first considered overall performance across the three block types. Proportions of correct trials were subjected to a one-way ANOVA with the variable of block (guess-and-learn block, test-and-learn block, reward block). The main effect of block was significant, $F(2, 68) = 374.9$, $p < .001$. Whereas performance was close to chance (30.7% ± 1.2%)[3] in the guess-and-learn block, it improved in the test-and-learn block (69.6% ± 2.5%) and the reward block (86.0% ± 2.0%). An important precondition of our analysis is that performance in the guess-and-learn block is entirely due to guessing because, otherwise, feedback-related brain activity could be biased by outcome expectancy (e.g., Hajcak et al., 2005). Unfortunately, performance in the first block was slightly better (30.7% ± 1.2%) than chance level (25%). Closer inspection of the data revealed that some Swahili probes were guessed correctly more frequently than others, which could reflect that some Swahili words contained subtle cues regarding their German meaning. To control for these outliers, we calculated, for each Swahili probe, the number of participants with correct guesses for this word (mean, 36.5%) and excluded all words for which this frequency was more than two standard deviations above or below the mean. In this way, 10 from 240 words were excluded (about 4% of the trials). A further cue that might have helped participants to determine the correct response without guessing is knowledge about the distractor. Because one or two distractors on each trial were taken from the current set of targets, it is possible that on some trials, two distractors were presented that were already targets in the same block and, thus, could easily be ruled out. When we excluded these trials (about 15% of the trials), performance in the guess-and-learn block nearly dropped to chance level (mean, 27.9%). As a consequence, we excluded these items from all further behavioral and ERP analyses. Note that neither of these exclusion procedures changed the results reported below qualitatively. They rather served to rule out the possibility that performance in the guess-and-learn block was not due to guessing.

In a further analysis, we investigated how efficiently participants learned from positive and negative feedback provided in the guess-and-learn block. To this end, we analyzed performance in the test-and-learn block as a function of guessing accuracy in the guess-and-learn block. A one-way ANOVA with the variable of guessing accuracy (correct, incorrect) revealed that performance in the test-and-learn block was improved if participants guessed correctly (73.6% ± 2.7%), as compared with guessing incorrectly (66.7% ± 2.7%), in the guess-and-learn block, $F(1, 34) = 12.8$, $p < .01$. However, although incorrectly guessed items were associated with an impaired performance in the test-and-learn block, this performance was still clearly above chance, $t(35) = 26.2$, $p < .001$, suggesting that participants made use of negative, corrective feedback to learn.

Positive versus negative feedback in feedback-learning blocks

In a next stage, we conducted an exploratory analysis to identify feedback-related ERPs in our paradigm. To this end, we considered feedback-locked waveforms elicited by positive and negative feedback in the guess-and-learn block. Figure 3 shows data from electrodes FCz (Fig. 3a) and Pz (Fig. 3b), which are representative for anterior and posterior activity in our data. Starting at about 200 ms after feedback onset, waveforms differed between positive and negative feedback. To quantify these effects, we calculated mean amplitudes for a series of time windows (I–IV) that were chosen to capture components of interest on the basis of visual inspection of the waveforms. Figure 3c shows the spatial distribution of amplitude differences between

[3] The second value represents the standard error of the mean.

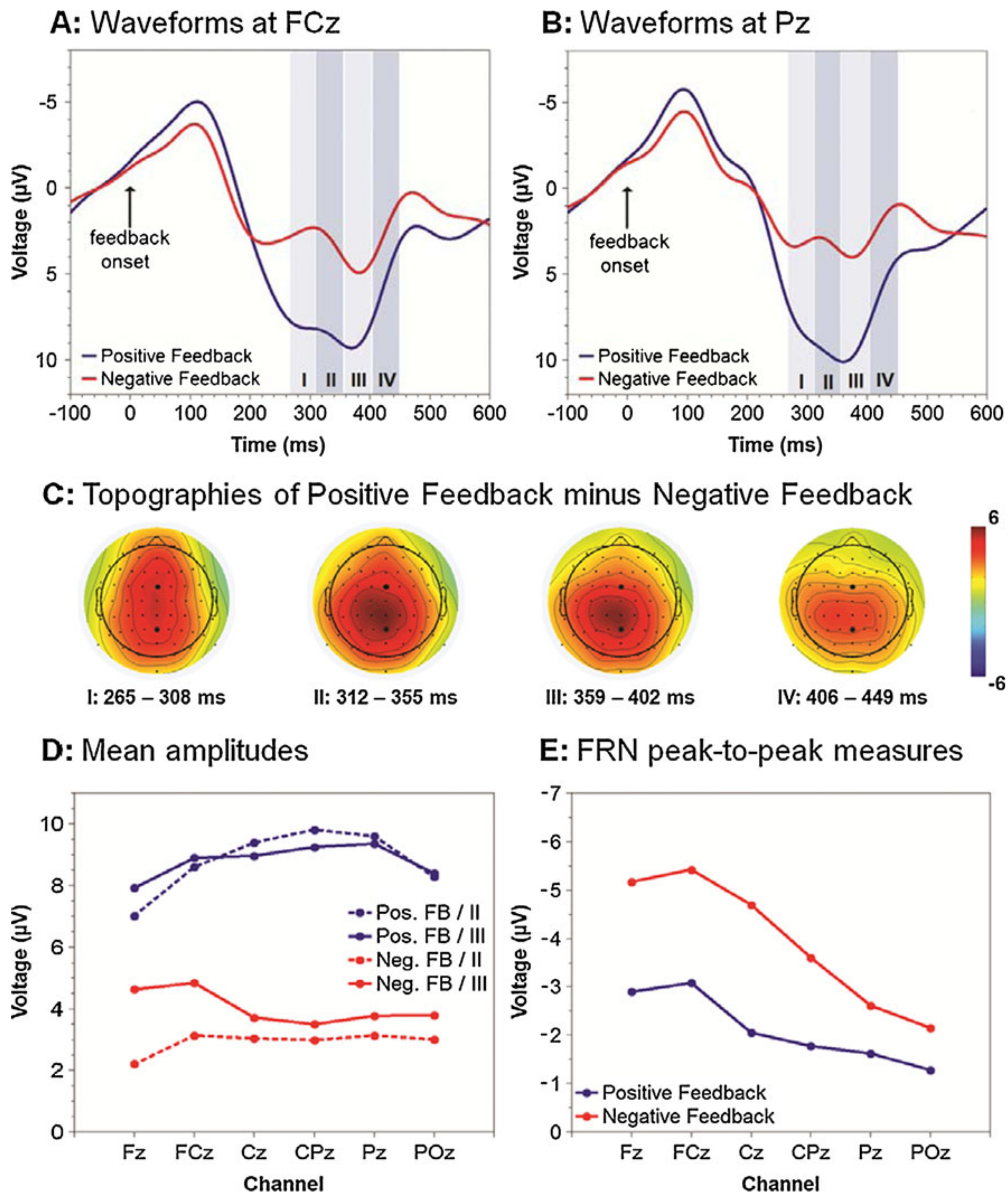

Fig. 3 Comparison of trials with positive and negative feedback in the guess-and-learn block. **a**, **b** Waveforms at electrodes FCz and Pz. **c** Spatial distribution of the differences between positive feedback trials and negative feedback trials in time windows I–IV. Large points indicate positions of electrodes FCz (upper point) and Pz (lower point). **d** Mean amplitudes across main midline electrodes for time windows II and III. **e** Peak-to-peak measures of the FRN across main midline electrodes. Pos. FB = positive feedback. Neg. FB = negative feedback

positive and negative feedback in these time windows. Whereas a strong difference was obtained across all time windows at posterior electrodes, anterior electrodes showed a difference mainly for earlier time windows. The posterior difference most likely reflects a feedback-locked P300 that is more positive for positive feedback than for negative feedback. In contrast, the early anterior difference presumably reflects an FRN that is more negative for negative feedback than for positive feedback.

To corroborate these observations statistically, we analyzed mean amplitudes for each feedback type (positive, negative) in the four time windows (I–IV) at two electrodes (FCz, Pz). Only effects involving the variable of feedback type and effects not qualified by higher-order interactions are reported. Because the three-way ANOVA revealed a significant three-way interaction, $F(3, 33) = 4.46$, $p < .01$, we continued with analyzing the data separately for each electrode. A two-way ANOVA with the variables of time window and feedback type on amplitudes at electrode FCz revealed a significant interaction, $F(3, 33) = 3.45$, $p < .05$, indicating that the difference between positive and negative feedback (presumably representing the FRN) was largest in the two earlier time windows and then decreased in the later time windows (I, 5.25 ± 0.88 μV; II, 5.47 ± 0.94 μV; III, 4.05 ± 0.78 μV; IV, 3.01 ± 0.76 μV). The same ANOVA on amplitudes at electrode Pz also revealed a significant interaction, $F(3, 33) = 3.00$, $p < .05$, indicating that the difference between positive and negative feedback (presumably representing the P300) was largest in the intermediate time windows but was smaller in the early and late time windows (I, 4.78 ± 0.78 μV; II, 6.47 ± 0.93 μV; III, 5.59 ± 0.74 μV; IV, 3.90 ± 0.69 μV). Note that the difference between positive and negative feedback was significant for all time windows at each electrode (all ps < .001).

Although the time course and topography of the early anterior difference between positive and negative feedback suggest that this effect corresponds to an FRN, an unequivocal interpretation of this effect is difficult because it strongly overlaps with the more posterior P300. To quantify the FRN independently of the P300, we applied a peak-to-peak analysis (Yeung & Sanfey, 2004). Figure 3e shows the distributions of the resulting FRN amplitudes across the main midline electrodes. The absolute FRN amplitude, as well as the difference between positive and negative feedback, is maximal at anterior electrodes, which is typical for an FRN. This receives support from a two-way ANOVA with the variables of feedback type and electrode, conducted on data from two representative electrodes (Pz, FCz). The interaction reached significance, $F(1, 35) = 5.78$, $p < .05$, indicating that a substantial difference between positive and negative feedback was obtained only for electrode FCz (2.34 ± 0.53 μV), $t(35) = 4.38$, $p < .001$, but not for electrode Pz (0.99 ± 0.63 μV), $t(35) = 1.58$, $p = .12$.

In addition to the feedback-locked P300 and the FRN, our data indicate that there is a third feedback-related ERP. The waveform at electrode FCz (Fig. 3a) reveals a positive peak for negative feedback in the third time window, which is smaller at electrode Pz (Fig. 3b) and which resembles the previously reported frontal positivity. To demonstrate that this effect represents an anterior positive peak and not only the peak of the posterior P300 coinciding with the tail of the FRN, Fig. 3d provides mean amplitudes at the main midline amplitudes separately for each feedback type from the third time window (and from the second time window for comparison). Mean amplitudes for negative feedback are maximal at electrode FCz, supporting the notion that this effect represents an anterior rather than a posterior peak. For positive feedback, this peak either is absent or is masked by the strong feedback-locked P300 peaking in the same time window. This conclusion receives support from a two-way ANOVA with the variables of feedback type and electrode, again conducted for two representative electrodes (Pz, FCz) for the third time window. A significant interaction, $F(1, 35) = 8.85$, $p < .01$, suggested that mean amplitudes were larger at electrode FCz than at electrode Pz for negative feedback (FCz: 1.07 ± 0.39 μV), $t(35) = 2.73$, $p < .01$, but not for positive feedback (Pz: −0.46 ± 0.50 μV), $t(35) = 0.93$, $p = .36$.

Taken together, our first analysis identified three components related to feedback processing in our paradigm. We found a feedback-locked P300 and an FRN that differed between positive and negative feedback. Although these components overlap in time, they can be clearly dissociated with respect to their spatial distribution and time course. Whereas the feedback-locked P300 peaks in the second and third time windows at posterior electrodes, the FRN peaks in the first and second time windows at anterior electrodes. Moreover, clear evidence for an FRN is also provided by a peak-to-peak analysis, which revealed the typical profile of an FRN with a peak at electrode FCz that is larger for negative than for positive feedback. Finally, we also identified a frontal positivity that seems to be larger for negative feedback. This component is difficult to separate from the P300, because it peaks at approximately the same time. However, the observation that it peaks at anterior electrodes (Fig. 3d) suggests that it is not simply a side effect of the parietal P300. Given that the frontal positivity is masked by the P300 for positive feedback, it is unclear whether it is really related to feedback processing. The following analyses will reveal whether this is indeed the case.

Predictors of successful learning in feedback-learning blocks

So far, the analyses of feedback-related ERPs in our multiple-choice task revealed three components that have previously been reported in the literature on feedback processing: a feedback-locked P300 that was larger for positive than for negative feedback, an FRN that was larger for negative than for positive feedback, and a frontal positivity for negative feedback that was either absent or masked by the P300 for positive feedback. In the following, we investigated which of these components is predictive for successful learning from negative, corrective feedback. To this end, we analyzed feedback-related ERPs for negative feedback trials in the guess-and-learn block as a function of whether a given item led to a correct or an incorrect response in the test-and-learn block. Only items with negative feedback in the guess-and-learn blocks and positive feedback in the test block (E–C items) and items with negative feedback in both blocks (E–E) were included. Learning from positive feedback was not further considered, because only very few items were associated with positive feedback in the guess-and-learn blocks and negative feedback in the test blocks.

Figure 4a, b shows waveforms for each trial type at electrodes Pz and FCz, and Fig. 4c shows the spatial distribution of the difference between E–C items and E–E items for the same four time windows used in the previous analyses. From visual inspection of these data, three conclusions can be derived: First, E–C items are associated with a larger feedback-locked P300, as indicated by a larger positivity across all time windows at posterior electrodes.

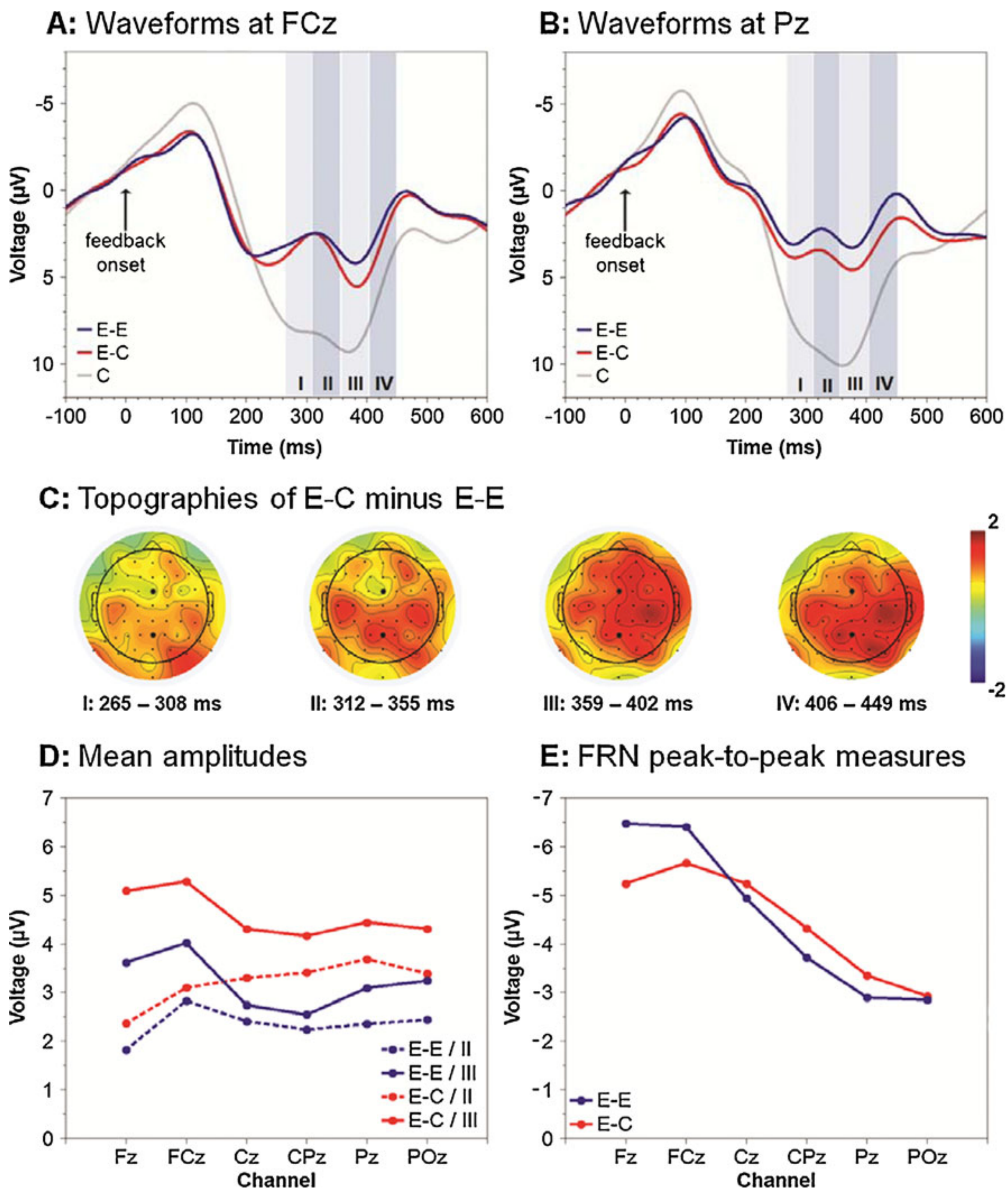

Fig. 4 Comparison of E–E items and E–C items in guess-and-learn blocks. **a**, **b** Waveforms at electrodes FCz and Pz. **c** Spatial distribution of the differences between E–C items and E–E items in time windows I–IV. Large points indicate positions of electrodes FCz (upper point) and Pz (lower point). **d** Mean amplitudes across main midline electrodes for time windows II and III. **e** Peak-to-peak measures of the FRN across main midline electrodes. E–E = incorrect responses in guess-and-learn block and test-and-learn block. E–C = incorrect response in guess-and-learn block but correct response in test-and-learn block

Second, E–C items and E–E items seem not to differ with respect to the FRN, as indicated by the absence of a difference at anterior electrodes in early time windows. Third, E–C items are associated with a larger frontal positivity, as indicated by a larger positivity in late time windows at anterior electrodes.

To corroborate these observations statistically, we analyzed mean amplitudes for each item type (E–C, E–E) at two representative electrodes (FCz, Pz). In contrast to the initial exploratory analyses, we included only the two time windows (II, III) for which the P300 and frontal positivity were maximal. Only effects involving the variable item type and only effects not qualified by higher-order interactions are reported. The three-way ANOVA revealed a significant three-way interaction, $F(1, 35) = 6.35$, $p < .05$. A subsequent two-way ANOVA with the variables of time window and item type at electrode Pz revealed a significant effect of item type, $F(1, 35) = 5.78$, $p < .05$, but no interaction ($F < 1$), indicating that the posterior difference between E–C and E–E (presumably representing the P300) was equally large in both time windows (1.34 ± 0.56 μV). In contrast, the same ANOVA at electrode FCz revealed a significant interaction, $F(1, 35) = 5.97$, $p < .05$, indicating that the anterior difference between E–C and E–E items (presumably representing the frontal positivity) was larger in the third time window (1.27 ± 0.72 μV) than in the second time window (0.27 ± 0.67 μV). The same results were obtained when data from electrode Fz, rather than from electrode FCz, were included.

To demonstrate that the positive peak in the third time window reflects a frontal positivity, Fig. 4d provides mean amplitudes for each item type across the main midline electrodes in the third time window (and the second time window for comparison). The figure shows that in the third time window, activity is maximal at electrode FCz, indicating a frontal peak. Moreover, whereas the difference between item types in the second time window is restricted to posterior electrodes (reflecting the P300 effect), the same difference in the third time window is additionally obtained at anterior electrodes. These observations receive support from statistical analyses of mean amplitudes at two representative electrodes (Pz, FCz). On the basis of the significant three-way interaction reported above, we now considered two-way ANOVA with the variables of electrode and item type. For the second time window, a significant interaction was obtained, $F(1, 35) = 5.24$, $p < .05$, indicating that the difference between E–C and E–E items was larger at electrode Pz (1.33 ± 0.56 μV) than at electrode FCz (0.27 ± 0.67 μV). For the third time window, however, the same ANOVA revealed only significant main effects of item type, $F(1, 35) = 4.21$, $p < .05$, and electrode, $F(1, 35) = 4.17$, $p < .05$, indicating that a difference for E–C and E–E items was present not only for electrodes Pz (1.35 ± 0.62 μV), but also for electrode FCz (1.27 ± 0.72 μV).

Finally, to corroborate the initial conclusion that item type did not influence the FRN, we analyzed peak-to-peak amplitudes. Figure 4e shows the distribution of FRN amplitudes across the main midline electrodes. A two-way ANOVA with the variables of item type and electrode, again conducted on data from two representative electrodes (Pz, FCz), revealed a significant effect of electrode, $F(1, 35) = 30.1$, $p < .001$, indicating larger amplitudes at electrode FCz (6.03 ± 0.62 μV) than at electrode Pz (3.12 ± 0.40 μV). Moreover, we obtained a significant interaction, $F(1, 35) = 4.42$, $p < .05$, reflecting that amplitudes at electrode Pz were larger for E–C items (3.35 ± 0.54 μV) than for E–E items (2.89 ± 0.44 μV), whereas amplitudes at electrode FCz were even larger for E–E items (6.41 ± 0.69 μV) than for E–C items (5.66 ± 0.70 μV), although the latter difference was not significant, $t(35) = 1.23$, $p = .23$. However, when the same analysis was computed using data from electrode Fz (rather than electrode FCz, for which the overall FRN was maximal), not only was the interaction between item type and electrode significant, $F(1, 35) = 6.53$, $p < .05$, but also amplitudes at electrode Fz were significantly larger for E–E items (6.47 ± 0.62 μV) than for E–C items (5.24 ± 0.69 μV), $t(35) = 2.06$, $p < .05$.

Taken together, these analyses suggest that successful learning from feedback is related to the amplitude of the feedback-locked P300, as well as to the amplitude of the frontal positivity. In contrast, although our feedback stimuli elicited a clear FRN, an increased FRN was not associated with more successful learning from feedback. Indeed, a larger peak-to-peak amplitude of the FRN at a frontal electrode (Fz) was even associated with impaired learning.

Discussion

The goal of the present study was to investigate which processes are crucially involved when participants learn from corrective feedback in a multiple-choice test. Multiple-choice tests combine characteristics of memory tasks and decision-making tasks. As in decision-making tasks, participants make a decision under uncertainty, followed by feedback about the correctness of their response. As in memory tasks, however, this feedback also includes information about the correct response. Whereas feedback in decision-making tasks has been viewed as a reinforcer (Frank & Claus, 2006; Holroyd & Coles, 2002), feedback in memory tasks has been viewed as a source of information for learning (Bangert-Drowns et al., 1991; Kulhavy, 1977). To investigate which of these processes are involved in learning from feedback in multiple-choice testing, we made use of the fact that several ERPs have been identified that are related to specific aspects of memory encoding and feedback processing. Our goal was not only to reveal which of these ERPs are involved in learning from feedback in multiple-choice testing; we also examined which of these ERPs are predictive of successful learning from feedback in this task.

In a feedback-learning condition, participants had to learn the correct response by guessing and then evaluating corrective feedback in an initial guess-and-learn block. Learning success was then tested in a second test-and-learn block and a final reward block. Behavioral data indicated that participants were able to learn from feedback. Performance was at chance level in the guess-and-learn block but reached nearly perfect accuracy in the final reward block. Interestingly, items that were incorrectly guessed in the guess-and-learn block were associated with impaired performance in the subsequent block. This is a frequently obtained finding (Kulhavy, 1977) and could reflect error perseveration due to the incidental encoding of incorrect information (Steinhauser, 2010; Steinhauser & Hübner, 2006).

A first analysis addressed the question of which feedback-locked ERP components were sensitive to feedback valence in the guess-and-learn blocks. We found the typical cascade of feedback-related ERP components known from research on decision-making tasks. Feedback elicited a typical fronto-central FRN that was larger for negative feedback than for positive feedback, followed by a parietal P300 that was larger for positive feedback than for negative feedback. These data suggest that feedback in multiple-choice testing triggers processes very similar to those in other decision-making tasks, including reinforcement learning (Holroyd & Coles, 2002) and feedback processing in working memory (Yeung & Sanfey, 2004). This is not surprising given that the decision required in the present paradigm is comparable to decisions used in research on decision making (e.g., deciding which Chinese character is rewarded in a specific context; see, e.g., Frank & Claus, 2006). In addition, we identified an early frontal positivity following negative feedback that strongly resembled the corresponding component found by Butterfield and Mangels (2003).

In a second stage, we investigated which of these components were predictive of successful learning. To achieve this, we compared feedback-locked ERPs between initially incorrect items from which participants successfully learned (E–C items) and those from which learning failed (E–E items)—a method that has previously been used to reveal ERP correlates of memory encoding (e.g., Karis et al., 1984; Paller et al., 1987). Our results indicated that several components predicted whether an item was later associated with a correct response or not. On the one hand, the amplitude of the feedback-locked P300 was larger in E–C items than in E–E items. On the other hand, the amplitude of the early frontal positivity was increased in E–C items, as compared with E–E items. In contrast, although a clear FRN was obtained, it was not positively correlated with learning success. When peak-to-peak amplitudes were considered, a larger FRN was even obtained for E–E items than for E–C items. In the following, we discuss the implications of these results for the question of which aspects of feedback processing are responsible for successful learning in our task.

The feedback-locked P300

The P300 has been shown to play an important role both in feedback processing and in memory encoding. In decision-making tasks, the feedback-locked P300 has been shown to vary with reward expectancies (Hajcak et al., 2005; Hajcak et al., 2006) and reward magnitude (Yeung & Sanfey, 2004), but not always with feedback valence (Yeung & Sanfey, 2004; but see Bellebaum & Daum, 2008; Bellebaum, Polezzi, & Daum, 2010; Hajcak, Moser, Holroyd, & Simons, 2007; van der Helden et al., 2010), which led to the idea that it is related to the update of outcome expectancies. In memory tasks, the P300 was related to learning success mainly under conditions where rote learning, rather than elaborate processing, was involved (Fabiani et al., 1990). Both findings are consistent with the idea that the P300 is related to the updating of working memory and, hence, to explicit encoding of new information (Donchin, 1981; Polich, 2007; Polich & Kok, 1995).

The present results integrate findings from literature on feedback processing and memory. First, we showed that the P300 is directly related to the processing of feedback about the correct response. This is suggested by the finding that positive feedback is associated with a larger P300 than is negative feedback (e.g., Hajcak et al., 2007). Second, the P300 for negative feedback was predictive of learning from feedback, replicating the dm effect (Paller et al., 1987) in the context of feedback processing. Together, these findings suggest that the P300 reflects a learning process that integrates both feedback about the initial response and information about the correct response.

To account for these results, we propose that the P300 to corrective feedback represents a learning process triggered by a feedback-based evaluation of the initial response in working memory. This account is based on the idea that learning from feedback is particularly fast and efficient if feedback confirms that a response that is already held in working memory is correct. In this case, an immediate update of working memory is triggered in which the correct response is linked to the probe and further context information that facilitates retrieval. The P300 could represent either working memory update itself or, alternatively, the decision process by which the feedback is evaluated (for a similar idea in the context of error detection, see Steinhauser & Yeung, 2010). Because participants normally hold only the selected response in working memory, fast learning indicated by a large P300 is obtained mainly for positive feedback trials. This can also account for the observation that performance is improved for items that were associated with a correct guess in the initial block. Under some conditions, however, fast learning can also occur for negative feedback trials—for example, because participants hold more than one response in working memory, or because participants decide for one answer but press another response. In this case, there is an increased P300 also for negative items, and this explains why the P300 is predictive of successful learning from negative feedback.

The frontal positivity

A second component that predicted successful learning from feedback was the early frontal positivity that immediately succeeded the P300. Butterfield and Mangels (2003) found a similar component following successful learning from corrective feedback in a semantic retrieval task. Because the frontal positivity varied with the expectedness of negative feedback, they assumed a functional relationship between this component and the P3a or novelty P3—a component of the stimulus-locked ERP reflecting an attentional orienting response to a novel or unexpected stimulus (Simons et al., 2001). Whereas the P3a typically precedes the posterior P300 (or P3b), this order seems to be reversed in the present data. This might reflect that the feedback-locked P300 is related to a fast learning process, whereas the frontal positivity represents attentional orienting that precedes a slower, more elaborate learning following negative feedback. If this interpretation is valid, a reduced frontal positivity in E–E items could reflect that a lack of attentional orienting has increased the probability that learning from feedback failed.

The finding that learning from corrective feedback is predicted by a frontal positivity resembles the results from the literature on memory processing that reports a late frontal positivity that predicts later memory retrieval (Fabiani et al., 1990; Kim et al., 2009; Mangels et al., 2001; Weyerts et al., 1997). Although the attentional orienting response represented by the early frontal positivity is presumably related to later—perhaps more elaborate—memory processing, it is unlikely that both phenomena represent identical mechanisms. The present frontal positivity occurs in the time range of the P300, whereas the late frontal positivity typically starts at 600 ms or later. The fact that we did not obtain a late frontal positivity that predicts learning success might be related to the present stimuli. Because our Swahili words convey no meaning by themselves, it is difficult to improve learning by more elaborate processing.

The FRN

Finally, we obtained an FRN that was larger for negative feedback trials than for positive feedback trials. Given that the FRN has been assumed to reflect reinforcement learning (Holroyd & Coles, 2002), this result clearly demonstrates that feedback in the present task is processed as a reinforcer. However, the amplitude of the FRN was not positively correlated with learning success. Rather, larger FRN amplitudes were even associated with impaired learning, although this result was obtained only when peak-to-peak amplitudes were considered. These results could reflect that reinforcement learning is an automatic process, which cannot be prevented but which does not improve learning success in the present task, presumably because learning requires explicit memory processes. The increased FRN for E–E items could reflect that on some trials, participants adopted a more reinforcement-related strategy at the cost of explicit learning (e.g., by attending more strongly to feedback valence than to information about the correct response), which enhanced the FRN on these trials while reducing the probability that an item was learned.

These results are consistent with a recent finding from Mangels et al. (2011), who investigated the relation between feedback-related brain activity in a complex math test with multiple-choice items and the participants' decision to voluntarily engage in further learning following negative feedback. They found that, at least in a stereotype threat condition, a strong FRN implied that participants were less willing to review the correct solution of a math problem. These and the present results are in accord with the common assumption made by dual-process theories that tasks typically involve both implicit and explicit processes and that performance sometimes reflects the one or the other (Ashby et al., 1998; Frank & Claus, 2006). They further highlight the role of strategies for learning in multiple-choice testing and illustrate how these strategies are reflected by ERP components.

Conclusion

Taken together, the present results suggest that two feedback-locked ERPs predict successful learning from corrective feedback in multiple-choice testing: the feedback-locked P300 and the early frontal positivity. We suggest that these ERPs are related to two different stages of learning. The P300 reflects a fast learning process based on working memory processes. In contrast, the frontal positivity reflects an attentional orienting response that precedes slower learning of correct response information. Finally, we obtained an FRN, which, however, was not positively correlated but, rather, negatively correlated with learning success. This finding suggests that feedback in multiple-choice testing is processed as a reinforcer, although reinforcement learning can even have detrimental consequences for future performance.

Author Note This research was supported by a grant to Marco Steinhauser from the Excellence Initiative of the University of Konstanz as part of the Center for Psychoeconomics. We are grateful to Lisa Kübler for assistance in conducting the experiments.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–481.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213–238.

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience, 27,* 1823–1835.

Bellebaum, C., Polezzi, D., & Daum, I. (2010). It is less than you expected: The feedback-related negativity reflects violations of reward magnitude expectations. *Neuropsychologia, 48*, 3343–3350.

Burnage, G. (1990). *CELEX: A guide for users*. Nijmegen: CELEX.

Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604–616.

Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research, 17,* 793–817.

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage, 49,* 3198–3209.

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience, 23,* 936–946.

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience, 27,* 371–378.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134,* 9–21.

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

Donchin, E. (1981). Presidential address, 1980. Surprise! .. Surprise? *Psychophysiology, 18,* 493–513.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology, 23,* 298–308.

Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a Von Restorff paradigm. *Electroencephalography and Clinical Neurophysiology, 75,* 22–35.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Reviews, 113,* 300–326.

Friedman, D., Nessler, D., & Johnson, R. (2007). Memory encoding and retrieval in the aging brain. *Clinical EEG and Neuroscience, 38,* 2–7.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295,* 2279–2282.

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology, 42,* 161–170.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology, 71,* 148–154.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology, 44,* 905–912.

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

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology, 16,* 177–216.

Kim, A. S. N., Vallesi, A., Picton, T. W., & Tulving, E. (2009). Cognitive association formation in episodic memory: Evidence from event-related potentials. *Neuropsychologia, 47,* 3162–3173.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47,* 211–232.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science, 14,* 47–53.

Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social Cognitive and Affective Neuroscience, 1,* 75–86.

Mangels, J. A., Good, C., Whiteman, R. C., Maniscalco, B., & Dweck, C. S. (2011). Emotion blocks the path to learning under stereotype threat. *Social Cognitive and Affective Neuroscience, 6,* 1–12.

Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11,* 77–95.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a 'generic' neural system for error detection. *Journal of Cognitive Neuroscience, 9,* 788–798.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology, 67,* 360–371.

Philiastides, M. G., Biele, G., Vavatzanidis, N., Kazzer, P., & Heekeren, H. R. (2010). Temporal dynamics of prediction error processing during reward-based decision making. *NeuroImage, 53,* 221–232.

Polich, J. (2007). Updating P300: An integrative theory of the P3a and P3b. *Clinical Neurophysiology, 118,* 2128–2148.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology, 41,* 103–146.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330,* 335.

Roediger, H. L., III, & Karpicke, J. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

San Martin, R., Manes, F., Hurtado, E., Isla, P., & Ibanez, A. (2010). Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *NeuroImage, 51,* 1194–1204.

Simons, R. F., Graham, F. K., Miles, M. A., & Chen, X. (2001). On the relationship of the P3a and the Novelty-P3. *Biological Psychology, 56,* 207–218.

Sommer, W., Heinz, A., Leuthold, H., Matt, J., & Schweinberger, S. (1995). Metamemory, distinctiveness, and event-related potentials in recognition memory for faces. *Memory & Cognition, 23,* 1–11.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30,* 641–656.

Steinhauser, M. (2010). How to correct a task error: Task-switch effects following different types of error correction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1028–1035.

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

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *Journal of Neuroscience, 30,* 15643–15653.

van der Helden, J., Boksem, M. A. S., & Blom, J. H. G. (2010). The importance of failure: Feedback-related negativity predicts motor learning efficiency. *Cerebral Cortex, 20,* 1596–1603.

Weyerts, H., Tendolkar, I., Smid, H. G. O. M., & Heinze, H. J. (1997). ERPs to encoding and recognition in two different inter-item association tasks. *NeuroReport, 8,* 1583–1588.

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex, 15,* 535–544.

Yeung, N., & Sanfey, A. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience, 24,* 6258–6264.