It Was Not MY Fault: Event-Related Brain Potentials in Active and Observational Learning from Feedback

Christian Bellebaum, Stefan Kobza, Stefan Thiele and Irene Daum

Institute of Cognitive Neuroscience, Department of Neuropsychology, Ruhr-University of Bochum, Universitaetsstrasse 150, D-44780 Bochum, Germany

Address correspondence to Dr Christian Bellebaum. Email: christian.bellebaum@rub.de.

Performance feedback during learning is accompanied by a negative event-related potentials (ERP) component, the feedback-related negativity (FRN), which codes a reward prediction error. An open issue relates to the coding of feedback stimuli in observational learning. The present study aimed to determine differences in the neural processing of feedback in active and observational learners in a between-subjects design. By choosing between different stimuli, 15 active learners could learn a rule determining the probability of monetary reward. Each of the 15 observers was yoked to the performance of one active learner. In test trials, observers could prove whether they had gained insight into the rule. Although both groups learned at a comparable rate, FRN amplitudes following negative feedback were significantly reduced in observational relative to active learners, whereas there was no difference for the FRN in response to positive feedback. Additionally, between-group differences were already observed in the time window preceding the FRN, between 150 and 220 ms after feedback onset. The processing of feedback stimuli thus depends upon the direct relevance for one's own action planning. The FRN as an error signal indicating the need for behavioral adaptation appears to be especially relevant, if negative feedback is linked to agency.

Successful adaptation to the environment requires the learning of associative links between actions and their consequences, with reward leading to enhanced and negative outcomes to reduced frequency of a behavior. As shown in monkeys, the mesencephalic dopamine system codes errors in reward prediction; the firing rates of dopaminergic neurons increase for unexpected rewards, and they are attenuated if an outcome is worse than expected (Schultz 2000). In their reinforcement-learning theory, Holroyd and Coles (2002) proposed that this dopaminergic prediction error signal is projected to the anterior cingulate cortex (ACC) from the basal ganglia to guide future selection of actions based on their outcomes. Within the basal ganglia, the ventral striatum including the nucleus accumbens is particularly involved in feedback processing (see Haber and Knutson 2010), and both the striatum and the ACC appear to play an important role in learning from feedback. In the monkey, striatal neurons integrate motor- and reward-related information (Schultz et al. 2000). In humans, striatal activations are consistently observed during feedback processing (e.g., Delgado et al. 2003; O’Doherty et al. 2004; Delgado 2007), and patients with selective striatal lesions are impaired in learning from feedback (Bellebaum et al. 2008).

The feedback-related negativity (FRN), a negative event-related potentials (ERPs) component evoked by the processing of response outcomes (Miltner et al. 1997; Holroyd and Coles 2002; see Nieuwenhuis et al. 2004), has been linked to the ACC (Gehring and Willoughby 2002; Bellebaum and Daum 2008). Subtraction of reward- from nonreward-related ERPs typically yields a negative deflection in the ERP, which peaks between 200 and 300 ms after feedback presentation (see Holroyd and Coles 2002; Nieuwenhuis et al. 2004). The FRN appears to be more pronounced, if negative feedback is unexpected or worse than expected, that is, when the prediction error is larger (Holroyd et al. 2003; Holroyd et al. 2004; Hajcak et al. 2007; Holroyd and Krigolson 2007; Bellebaum and Daum 2008; Holroyd et al. 2009), although this has not always been observed (Hajcak et al. 2005).

Despite a wealth of recent data, the nature and the functional implications of the FRN are still under debate. According to Holroyd and Coles (2002), the learning signal mirrored by the FRN is used to optimize future behavior of the acting subject. In this context, it is important to note that the FRN resembles another ERP component, the error-related negativity (ERN). The ERN is a response-locked signal that is typically observed when subjects commit errors in simple stimulus–response tasks (Falkenstein et al. 1990; Gehring et al. 1993; Dehaene et al. 1994). The close resemblance between FRN and ERN appears to suggest that both components are neural signatures of performance monitoring. The feedback-locked FRN reflects a learning signal for response-outcome associations, whereas the ERN signals performance errors related to already learned behaviors (Holroyd and Coles 2002; Frank et al. 2005). Associations between actions and their outcomes can, however, also be learned by observing the behavior of others. Learning by observation may save time and energy and it prevents exposure to dangerous situations.

In a seminal paper, van Schie et al. (2004) showed that the mere observation of errors yields a negative ERP component. Rather than resembling the error-related potentials typically found in active performers, the potential observed by van Schie et al. (2004) appears to be similar to the FRN: The peak latency is about 250 ms after error observation, and the negativity following observed errors does not present as a large peak but as a small deflection. However, the data base for the processing of performance “feedback” in an observation situation, in which feedback does not refer to the subject’s own performance but to the performance of another person, is as yet very sparse, and the findings are less clear. In a recent study, qualitatively similar FRN amplitudes were found in active and observation conditions (Yu and Zhou 2006): Amplitudes were higher following losses compared with gains in both conditions, which led Yu and Zhou (2006) to conclude that the neural mechanisms for active and observational learning from feedback are similar. This finding is surprising as negative
feedback has no direct implications for immediate behavioral adaptation in a purely observational condition, and there is thus no need for error signal processing. Although more recent studies reported reduced FRN amplitudes in observation conditions (Itagaki and Katayama 2008; Fukushima and Hiraki 2009), the overall pattern of ERPs was comparable in active responders and observers. One reason for the similarity of the signals might be related to the study design. In all previous studies examining feedback processing in an observation condition, subjects took turns with a “collaborator”—mostly on a trial-by-trial basis, and carryover effects may have prevented a clear separation of the processes involved in the active and the observation condition.

We hypothesized that the reward system is more strongly involved in feedback processing in active learners compared with subjects purely learning by observation and expected clear differences in the neural processing of monetary outcome stimuli. Focusing on the FRN, negative feedback was expected to yield larger amplitude FRNs in active compared with observational learners. Active and observational learning were examined in a between-subjects design to reduce potential between-trial carry-over effects with respect to learning and feedback coding. A variant of a previously described task was used, in which outcome frequencies are not predetermined but depend upon subjects’ insight into a reward-predicting rule (see Bellebaum and Daum 2008). It has been shown that FRN amplitude more reliably reflects a negative prediction error, if stimulus-response outcomes can actually be learned (Holroyd et al. 2009). We found that active learners and observers learned a rule determining reward probability equally well. With respect to the processing of rewarding and nonrewarding outcome stimuli, clear differences were observed between the groups in FRN amplitude as well as in the time window of a positive peak preceding the FRN.

Materials and Methods

Participants
Thirty-two healthy, right-handed subjects (18 female, 14 male) participated in this study. All subjects were students of the Ruhr-University of Bochum and had normal or corrected-to-normal vision. The data of 2 subjects had to be excluded from analysis because of technical data acquisition problems. The remaining subjects formed 2 groups of 15 subjects each, matched for age. The “active learners” (7 women and 8 men; mean age = 25.1 years, standard deviation [SD] = 4.3 years) performed a task which required learning a rule which predicted reward depending on response-outcome links and the “observers” (9 women and 6 men; mean age = 25.5, SD = 4.3 years) were asked to learn the rule on the basis of observing the outcomes (reward/nonreward) of the actions of another person. Each subject of the observer group was yoked to the behavior of an active learner (see below for details on the task).

An IQ estimate was obtained using the subtests “Picture Completion” and “Similarities” of a short version of the German Wechsler Adult Intelligence Scale (Dahl 1972). Mean IQ scores for the active and observer groups were 115.7 (SD = 8.7) and 111.8 (SD = 7.2) and did not differ significantly (P = 0.19). The study was approved by the Ethics Committee of the Faculty of Medicine of the Ruhr-University of Bochum, and all participants gave written informed consent.

The Active Learning Task
A variant of a previously administered learning task was used (Bellebaum and Daum 2008). On each trial, subjects had to guess the location where a 5-cent coin was hidden. If they found the coin, they had won it (rewarding outcome). If they did not find the coin, there was no gain (nonrewarding outcome). Subjects were instructed that the coin was hidden in 1 of 6 boxes. On each trial, following a fixation cross, 2 stimuli consisting of sets of 6 boxes were shown on the left and on the right side of the computer screen. In each set, different boxes were “preselected,” as indicated by red color (gray in Fig. 1). The subjects were told that the 2 sets on the left and right represented the same 6 boxes, the only difference being the pattern of preselected boxes. By pressing a left or right response button, subjects could choose the left or right subset, depending upon whether they thought that the 5-cent coin was hidden in one of the preselected red boxes in the left or the right set. Thus, subjects did not have to guess the location of the individual box, but they had to choose between sets of preselected boxes.

The stimuli remained on the screen until subjects chose one of the stimuli with a button press; the maximum duration was 2700 ms. As soon as a choice had been made, only the selected set was visible for 500 ms. After another 500-ms interval (black screen), the feedback (reward or nonreward) was displayed for 500 ms (see Fig. 1A for the sequence of events in the active learning task). The intertrial interval was 700 ms.

As illustrated in Figures 1 and 2, each set was arranged in 2 rows of 3 boxes each. The total number of preselected boxes was 2 or 4. Unknown to the subject, the 5-cent coin was always hidden in the lower row of a set, and reward probability was thus determined by the number of preselected boxes in the lower row. Figure 2 provides an overview of the different stimulus types and associated reward probabilities.

The learning task comprised 3 blocks of 220 trials each. After the first block, subjects were told to focus on the distribution of boxes, which served as a cue to figure out the rule determining reward probability. After completion of the experiment, they received the sum corresponding to their gains, that is, the rewards accumulated during the course of the task.

The Observational Learning Task
Subjects of the observational learning group were told that they would observe the performance of another person (“subject X”) during a learning task. They read the same instructions as the participants of the active learning group, being also told that these were the actual instructions of the person they were about to observe. The sequence of events on an individual trial was identical to the sequence in the active learning task, with the exception that the observers did not choose a subset of boxes themselves but were shown the choice of the observed subject. Importantly, each of the 15 observers was yoked to the performance of one of the 15 active subjects, such that all relevant experimental variables were perfectly matched in pairs of subjects. On each individual trial, observers saw the same types of stimuli and the response of the corresponding active learner. Active learners’ choices were indicated by a frame around the chosen box set that appeared after a delay matching the observed subject’s response time on that particular trial. To match the motor requirements of the active task, observers had to indicate the choice made by the observed subjects by pressing the left or right response button within a maximal response time of 1500 ms. Immediately following the observed response, the nonchosen stimulus disappeared and only the subset of boxes chosen by the observed subject remained on the screen, as in the active learning task. Finally, the outcome received by the active subject appeared (see Fig. 1B). If the observers pressed the wrong button (i.e., the one not pressed by the observed subject) or did not press a button at all, they were asked to respond correctly or faster, respectively.

After each block of 220 trials, the observers completed 24 active test trials. On each of these trials, subjects had to choose 1 out of 4 box sets of the type used in the learning task by pressing 1 of 4 different response buttons which were different from those buttons used to indicate the observed subject’s responses during the learning phase. On each test trial, one of the box sets was associated with a higher reward probability than the other stimuli. Subjects were instructed to choose the stimulus which they thought was related to the highest reward probability. No feedback was given on test trials, that is, subjects’ choices could only be based on their previous observations. The test
blocks were introduced (1) to assess whether observers had gained insight into the rule determining reward probability and (2) to enhance the observers’ motivation to learn from the response-outcome data of the person they observed. They were also told that they would be paid out the gains of the person they observed if they showed in the test trials that they had learned the reward-determining rule by observation. The outcomes thus had the same significance for both active learners and observers. As the active learners, observers were given a cue to attend to the spatial distribution of the boxes after completion of the test trials that followed the first block of trials.

Electroencephalography Recording

Subjects were comfortably seated approximately 70 cm in front of a computer monitor. The left and right CTRL keys of a computer keyboard were used as response keys. During performance of the learning task, electroencephalography (EEG) was recorded from 30 scalp sites according to the International 10–20 system with silver-silver chloride electrodes: F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, CZ, C4, T8, TP7, CP3, CPz, CP4, TP8, P3, Pz, P4, P8, PO7, PO3, POz, PO4, and PO8. Electrooculography (EOG) was recorded from the outer canthi of both eyes and above and below the left eye to monitor horizontal and vertical eye movements. Recordings were referenced to the algebraic average of the left and right mastoids. Stimulus timing was controlled by Presentation Software (Neurobehavioral Systems Inc.). Data were recorded with a sample rate of 500 Hz using a Neuroscan Synamps system (bandpass: 0.01–100 Hz) and the appropriate software. Impedances were kept below 10 kΩ.

Data Analysis

EEG Data

EEG and EOG data were analyzed off-line using the Brain Vision Analyzer Software Package. After applying a 0.1-Hz high-pass and a 40-Hz low-pass filter, an independent component analysis (ICA) was performed on single-subject EEG data (Lee et al. 1999). ICA yields an unmixing matrix, which decomposes the multichannel scalp EEG into a sum of temporally independent and spatially fixed components. The number of components matches the number of channels. Each resulting component is characterized by a time course of activation and a topographical map. In accordance with the procedure used in a previous study (Bellebaum and Daum 2008), each subject’s 30 components were screened for maps with a symmetric, frontally positive topography, which might represent eye movement and blink artifacts. These components were then removed from the raw data by performing an ICA back transformation. In most subjects, one such component was identified and removed. Only if the back-transformed data still contained numerous eye movement and blink artifacts, as indicated by visual inspection, a second component was removed. Then an automatic artifact detection technique that automatically excludes trials with data points exceeding an absolute amplitude value of 100 μV was applied to back-transformed data. The number of trials that had to be excluded from analysis was generally very low. In active learners, on average, 2.6% of trials were discarded (SD = 6.0%). In observational learners, the mean number of discarded trials amounted to 1.2% (SD = 2.4%). None of the subjects had to be excluded because of too many artifacts.

To analyze feedback-related ERPs, segments were created from 200 ms before to 800 ms after feedback presentation (reward or nonreward). The ERPs at electrode positions Fz, FCz, and Cz were pooled for FRN analysis as the FRN is maximal at these sites. Similar to procedures applied in previous studies (e.g., Holroyd et al. 2009) and in accordance with theoretical considerations (Luck 2005), we first analyzed the difference waves, subtracting ERPs following reward from the corresponding ERPs following nonreward, yielding separate difference waves for expected and unexpected (i.e., high and low probability) outcomes. In a first step, the peak of the difference wave was extracted for every subject and condition, defined as the maximum negative peak in the time window between 100 and 300 ms after feedback onset. To further explore time windows, in which the mean amplitude of the ERP differed significantly between active and observational learners, separate analyses of variance (ANOVA) were carried out for consecutive time bins of 10 ms between 100 and 400 ms after feedback onset were conducted (Rugg et al. 1995). The time windows of significant amplitude difference were then analyzed further by taking into account the original ERPs that is, the reward- and nonreward-related ERPs. In a last step, the amplitude of the FRN was defined as the maximum negative peak amplitude in the time window between 200 and 340 ms after onset of feedback presentation, relative to the positive peak amplitude between 150 ms after feedback onset and the latency of the negative peak.

Behavioral Data

Analysis of the behavioral data aimed to determine, if individual subjects learned to predict reward probabilities on individual trials. In the active

Figure 1. Trials and stimuli: Sequence of events on an individual trial for (A) active and (B) observational learning. In the active learning task, the chosen stimulus remained on the screen for 500 ms, whereas the alternative stimulus disappeared. In the observational learning task, a frame signaled the choice made by the observed subject. Then the chosen stimulus remained on the screen alone, as in the active learning task. To see whether the observed subject was rewarded or not, observers had to indicate the observed subject’s choice with a button press.

Figure 2. Types of stimuli and reward probabilities: For each reward probability and total number of preselected boxes, 3 different stimuli were used.
learners, the choice behavior in learning trials was analyzed. In 600 of the 660 trials, reward probabilities differed between the 2 box sets, enabling the subject to choose the alternative with the higher reward probability.

In the observers, the active test trials were analyzed. Each observer completed 3 blocks of 24 test trials each, that is, after each block of 220 observation trials. Although the number of 24 test trials is quite low, subjects had to choose between 4 alternatives on each trial, with only 1 stimulus being associated with a higher reward probability compared with the others. On average, simple guessing would lead to 6 correct responses out of 24, and above-chance-level performance can be reliably identified.

The Different Conditions
The conditions entering ERP analysis were determined by the active subjects’ choices and their outcomes. Only those trials were considered, in which subjects chose (or observed the choice of) stimuli with reward probabilities of 1/3 or 2/3 ("1/3 and 2/3 choices"). Both types of choices might yield reward or nonreward, leading to 4 possible combinations: high-probability reward (rewarded 2/3 choice), high-probability nonreward (nonrewarded 1/3 choice), low-probability reward (rewarded 1/3 choice), and low-probability nonreward (nonrewarded 2/3 choice). Subjects were expected to gain insight into reward probabilities during the course of the experiment. Therefore, 2 stages were considered separately—one before and one after subjects had learned the rule (see Results).

Statistical Analyses
Response accuracy in the 3 learning (active learners) or test phases (observers) was analyzed with a repeated-measures ANOVA involving the factor BLOCK (1-3). For the ERPs, nonreward–reward difference waves were analyzed in a first step (see above). By means of repeated-measures ANOVAs for consecutive time bins of 10 ms including the between-subjects factor GROUP (active learners vs. observers), time windows of significant amplitude difference between groups were identified. In the following, difference waves’ peak amplitudes as well as mean amplitudes and the FRN peak of the original ERPs were analyzed (see Results for details). For all statistical analyses, the level of significance was set to \( P < 0.05 \). The Greenhouse-Geisser correction to adjust the degrees of freedom was applied when the sphericity assumption was violated.

Results
Behavioral Data
Figure 3 illustrates the course of choice behavior of active subjects and observers. Both groups of subjects clearly acquired the reward-determining rule. More specifically, active learners showed a steep improvement early in block 2, after the cue to focus on the spatial distribution of preselected boxes had been given (Fig. 3A). On average, active learners chose the box set with higher reward probability ("correct responses") in 52% of trials in block 1 (SD = 6%), 82% in block 2 (SD = 11%), and 87% in block 3 (SD = 14%). Repeated-measures ANOVA revealed a significant increase in the number of correct responses (\( F_{1,120}^{2,19,460} = 82.848; P < 0.001 \)), with significant improvements from block 1 to block 2 (\( F(1) = -11.148; P < 0.001 \)) and from block 2 to block 3 (\( F(1) = -2.182; P = 0.047 \)).

Similarly, the course of the performance of the observers on test trials indicates that they gained insight into the reward-predicting rule (Fig. 3B). Overall, the number of correct responses on the test trials increased significantly across blocks (\( F_{1,139}^{15,949} = 20.457; P < 0.001 \)), with significant improvements from block 1 to block 2 (\( F(1) = -4.718; P < 0.001 \)), but no further improvement from block 2 to block 3 (\( P = 0.456 \)).

Overall, there were clear similarities in the course of learning in both groups. They gained insight into the rule during the second learning block, reaching maximum performance levels toward the end of the second block: As pointed out above, observers did not increase their performance level further after the second test block. Similarly, the performance levels of active learners were quantitatively similar in the second half of block 2 (89% correct responses; SD = 11%) and in block 3 (see above). To evaluate learning effects, only the trials in the first and third learning blocks were considered for ERP analysis, representing the pre- and postlearning phase, respectively.

Feedback-Related Potentials
Difference Waves
As was outlined in the Materials and Methods, the nonreward–reward difference waves were analyzed first. Difference waves for expected and unexpected outcomes in the pre- and postlearning phases for active and observational learners are illustrated in Figures 4A and B. An ANOVA on difference wave peak amplitudes comprising the factors PHASE (pre- vs. postlearning), PROBABILITY (high vs. low), and GROUP (active vs. observer) revealed main effects of GROUP (\( F_{1,28} = 6.648; P = 0.015 \)) and PHASE (\( F_{1,28} = 7.135; P = 0.012 \)), indicating larger amplitudes in active learners compared with observers and in the pre- compared with the postlearning phase. A trend for the factor PROBABILITY (\( P = 0.075 \)) suggests that difference wave peak amplitudes tended to be larger for unexpected outcomes. All interactions did not reach or approach significance (all \( P > 0.499 \)).

Visual inspection of the difference waves appears to suggest that the ERPs of both groups differ mainly in 2 time windows. To characterize these time windows in more detail, separate ANOVAs with the abovementioned factors were conducted for mean amplitudes in consecutive 10-ms time windows between 100 ms and 400 ms after feedback presentation. The analysis yielded significant group differences between 150 ms and 220 ms and between 280 ms and 330 ms (\( P < 0.05 \) for all comparisons).

Positive and Negative Feedback
In a second step, the neural responses to reward and nonreward were analyzed directly. In Figures 4A and B, grand average ERPs evoked by high- and low-probability positive and negative feedback in the pre- and postlearning phases are shown, separately for active learners and observers. Separate 4-way ANOVAs with the 3 factors defined above and the additional factor OUTCOME (reward vs. nonreward) were carried out for mean amplitude measures in the 2 time windows.
windows of significant group difference revealed by the difference wave analysis (see above). For the time window 150–220 ms after feedback presentation, significant main effects of GROUP \( (F_{1,28} = 7.740; P = 0.010) \), OUTCOME \( (F_{1,28} = 46.493; P < 0.001) \), and PHASE \( (F_{1,28} = 31.181; P < 0.001) \) emerged. Amplitudes were less positive for observers compared with active learners and for negative compared with positive feedback. Positive amplitudes were less pronounced in the post- compared with the prelearning phase. In addition, there was a significant interaction between OUTCOME and GROUP \( (F_{1,28} = 17.070; P < 0.001) \). Separate comparisons of amplitudes related to reward and nonreward in active and observational learners—across phases and probabilities—yielded significant nonreward – reward amplitude differences in both groups, which were more pronounced for active learners (active learners: \( t_{(14)} = -6.322; P < 0.001 \); observers: \( t_{(14)} = -2.687; P = 0.018 \)). All other main effects or interactions did not reach significance (all \( P > 0.050 \)).

Analysis of the 280- to 330-ms time window yielded similar results. Main effects for GROUP \( (F_{1,28} = 15.707; P < 0.001) \) and
PHASE ($F_{1,28} = 42.196; P < 0.001$) were again related to larger positive amplitudes in active learners relative to observers and in the prelearning phase relative to postlearning. A significant OUTCOME effect was not observed ($P = 0.468$), but a significant interaction between OUTCOME and GROUP ($F_{1,28} = 9.375; P = 0.005$) emerged. Separate follow-up $t$-tests for the 2 groups, as already conducted for the early time window, revealed significant amplitude differences between nonreward and reward for both active learners ($t(14) = -2.176; P < 0.047$) and observers ($t(14) = 2.382; P < 0.032$). However, for active learners, amplitudes were more negative (or less positive) for negative outcomes, whereas for observers positive outcomes were accompanied by more negative (less positive) amplitudes.

As can be seen in the original ERP traces in Figures 4A and B, the 280- to 330-ms time window does not comprise the negative peak reflecting the FRN in all groups and conditions. Because of these between-group or between-condition differences in FRN latencies, the analysis of mean amplitudes may not suffice to characterize the modulations of outcome processing in the present study. Therefore, an additional analysis of individual FRN peaks was conducted (see Methods for definition of FRN amplitude).

An ANOVA on FRN latencies yielded main effects of GROUP ($F_{1,28} = 34.149; P < 0.001$; latencies shorter in active learners), OUTCOME ($F_{1,28} = 23.943; P < 0.001$; latencies shorter for nonreward), and PHASE ($F_{1,28} = 7.178; P = 0.012$; latencies shorter for the prelearning phase; see Fig. 5). Furthermore, a significant interaction between the factors GROUP and OUTCOME emerged ($F_{1,28} = 5.025; P = 0.033$). In both active ($t(14) = -2.549; P = 0.023$) and observational learners ($t(14) = -4.177; P = 0.001$), FRN latencies for nonreward were shorter than for reward, but the latency difference was more pronounced in the observers. All remaining main effects or interactions did not reach significance (all $P > 0.212$).

The analysis of FRN amplitudes yielded a main effect of OUTCOME ($F_{1,28} = 24.795; P < .001$), indicating generally higher FRN amplitudes for reward compared with nonreward (see Fig. 6 for an illustration of peak amplitudes). A significant interaction between the factors OUTCOME and GROUP further shows that the reward - nonreward difference was modulated by the type of learning, active or observational ($F_{1,28} = 4.268; P = 0.048$). Follow-up $t$-tests revealed that active and observational learners did not differ in the FRN amplitude following reward ($t(28) = 0.049; P = 0.961$). However, FRN amplitudes following nonreward were significantly larger in active compared with observational learners ($t(28) = -2.545; P = 0.017$). Furthermore, a significant 3-way interaction between PHASE, OUTCOME, and PROBABILITY was observed. In follow-up $t$-tests, FRN amplitudes were compared directly between high- and low-probability outcomes, separately for reward and nonreward before and after learning had taken place. These comparisons did not yield any differences between high- and low-probability outcomes, apart from the comparison between unexpectedly and expectedly rewarded trials in the postlearning phase. For unexpected reward, the FRN tended to be larger compared with expected reward ($P = 0.095$, $P = 0.043$ tailed; for all remaining $t$-tests $P > 0.168$). None of the remaining main effects or interactions reached significance (all $P > 0.140$).

**Discussion**

Behavioral adaptation can be accomplished in different ways. The most straightforward way is learning from the consequences of one’s own actions, that is, increasing or reducing the frequency of behavior depending on the positive or negative outcome. Alternatively, adaptive behavior can be acquired by observing the outcome of the actions of others. The present study aimed to elucidate, whether the neuronal mechanisms underlying feedback processing during observational learning differ from those involved in active learning. In contrast to previous studies examining this issue (Yu and Zhou 2006; Itagaki and Katayama 2008; Fukushima and Hiraki 2009), a between-group design was applied, with one group learning actively from their own choices and the accompanying outcome, and the other group by observing responses and outcomes in others.

As revealed by difference wave analyses, the peak amplitude difference between nonreward and reward was significantly more pronounced in active learners. The largest difference between ERPs following nonrewarded and rewarded outcomes was, however, not observed in the time window of the FRN, which is typically associated with feedback processing, but
earlier, between about 200 and 220 ms in both active and observational learners. The comparison of difference waves' mean amplitudes in consecutive 10-ms time windows revealed group differences not only at this early stage of feedback processing (between 150 and 220 ms) but also between 280 and 330 ms. Inspection of the grand average ERP waveforms suggests that between-group differences in FRN latencies may have contributed to the difference wave deviations in the later time window. Indeed, FRN latencies were shorter in active learners and were also affected by the valence of the outcome. Therefore, an analysis of FRN peak amplitudes was conducted in addition to the mean amplitude analyses. Although active and observational learners did not differ with respect to reward-related FRN amplitudes, the FRN following nonreward was significantly larger in active learners compared with observers. Surprisingly, FRN amplitudes in both groups were not more pronounced in response to negative outcomes compared with positive outcomes, as was described by most previous studies (see Nieuwenhuis et al. 2004). This finding was obtained because the FRN was scored relative to the preceding positivity, which was also affected by outcome valence. A similar pattern of comparable FRN amplitudes for negative and positive feedback has also been observed in other studies applying the same scoring procedure (Frank et al. 2005; Oliveira et al. 2007). When the absolute peak values for the negative peak in the time window of the FRN were considered, the typical pattern of more negative FRN amplitudes for nonreward was observed in active, but not in observational, learners. The main focus of the present study was, however, on between-group differences in FRN amplitudes.

Previous studies showed that the magnitude of the negative prediction error is also coded by the FRN, with larger amplitudes the more unexpected the outcome (Hajcak et al. 2007; Holroyd and Krigolson 2007; Bellebaum and Daum 2008; Holroyd et al. 2009). In the present study, a significant 3-way interaction between OUTCOME, PROBABILITY, and PHASE suggested that FRN amplitude was affected by reward expectancy. This interaction was, however, caused by a modulation of reward-related FRN amplitudes in the postlearning phase.

Taken together, the analyses reported above provide evidence for differences in reward processing between active and observational learning. This finding is remarkable given the fact that active and observational learners did not differ with respect to the level of insight into reward probabilities. Thus, both types of learning were similarly successful but accompanied by differential patterns of ERPs.

The main findings of the present study appear to contrast with previous results on performance monitoring in active and observation conditions. Studies on error (van Schie et al. 2004) and feedback processing (Yu and Zhou 2006; Itagaki and Katayama 2008) emphasized the similarities in neural activation patterns between the monitoring of the subjects' own behavior and the behavior of an observed person. In the study by van Schie et al. (2004), subjects actively performed a flanker task, followed by a passive condition, in which another subject was observed performing the same task. In the active condition, performance errors elicited a typical ERN. Although time locked to an observed response, the error-locked negativity in the observation condition resembled an FRN with respect to latency and waveform. Along similar lines, larger FRN amplitudes for negative compared with positive outcomes were reported for both active and observation conditions of a gambling task with monetary feedback (Yu and Zhou 2006; Itagaki and Katayama 2008).

In the present study, clear differences in outcome processing were observed between active and observational learners. However, with respect to the FRN, the results appear to be compatible with what has been found in previous studies. Both Yu and Zhou (2006) and Itagaki and Katayama (2008) observed more pronounced FRN amplitudes for an active compared with an observation condition, but they either did not analyze the difference directly or did not focus on it in the discussion of their results. The differences between outcome processing in active and observation condition observed in the present study go beyond these previous findings. We could show that the first differences occur already in an earlier time window around 200 ms after feedback onset, when a positive peak precedes the FRN. Potts et al. (2006) proposed that the positive P2a constitutes an independent feedback-related ERP component (Potts et al. 1996), and the early difference between active learners and observers found in the present study appears to reflect a quantitative reduction in this early component in the observers.
Methodological differences may underlie the divergence of the current compared with previous results. All previous studies were based on a within-subjects design, examining the same subjects in an active and an observation condition. In the feedback processing studies, subjects took turns with the virtual observed person in choosing risky or nonrisky stimuli on a trial-by-trial basis, and they had to work out hypotheses on a possible rule underlying reward probabilities based on the outcomes of both active and observed responses (Yu and Zhou 2006; Itagaki and Katayama 2008; Fukushima and Hiraki 2009). With such a design, it is not unlikely that feedback stimuli for the observed person were processed in relation to representations of own responses, which were required on the preceding and the following trials, possibly leading to more comparable neural coding of outcomes in the active and observation conditions than in the present study. In the current between-group design, the 2 conditions were completely separated. Although observational learners needed to use the knowledge gained during observation, they had to apply it in a different context, that is, the test trials on which the response format was different (choice among 4 stimuli, different response buttons, no feedback). It is thus not likely that learning in observers was accompanied by the development of an association between specific responses and accompanying outcomes. In order to know the outcome of an observed action during learning, observers had to follow the choice made by the observed person by pressing the same button. They knew, however, that it was not their choice, which led to reward or nonreward.

The finding of reduced FRN amplitudes following negative feedback in observational learners is consistent with the assumptions of the reinforcement-learning theory of the FRN, outlined by Holroyd and Coles (2002). The FRN signals the need for behavioral change when an outcome is worse than predicted in active learning (Hajcak et al. 2007; Holroyd and Krigolson 2007; Bellebaum and Daum 2008; Holroyd et al. 2009; see Holroyd and Coles 2002). When negative feedback follows the action of another person, the signal is much weaker because the response leading to negative feedback was not committed by the observer him-/herself, and behavioral change or adaptation is thus not necessary. FRN amplitude appears to depend on the sense of agency for the behavior in question. Assuming that the neural mechanisms underlying active and observational feedback learning are dissociable, it might be surprising to find a quantitative reduction of the FRN for negative feedback in observational learning rather than a complete absence. In this context, it is interesting to note that the absolute FRN peak value for negative feedback did not differ from the value for positive feedback in the observers, which might even provide some support for qualitative differences in neural coding between the 2 types of learning. However, we decided for a peak-to-peak analysis to account for amplitude differences in the P200 time window, and this analysis yielded the above described quantitative difference. Further support for a reduced sense of agency being reflected in quantitative FRN amplitude reductions comes from studies in which reduced but significant FRN peak values for negative feedback were observed in the complete absence of actively performed or observed responses (Donkers and van Boxtel 2005; Donkers et al. 2005; Yeung et al. 2005; Potts et al. 2006). Yeung et al. (2005) directly compared one condition in which subjects experienced outcome stimuli as dependent on their own choice behavior with conditions in which subjects did not have a choice or did not have to respond at all. FRN amplitude decreased with the degree of active involvement of the subject. Decreased involvement meant, for example, that subjects rated the no-choice or no-response conditions as less interesting and had reduced affective responses to the outcomes in these tasks. It is unlikely that the observers in the present study showed a similar decrease in involvement. In contrast to the study by Yeung et al. (2005), the frequencies of different outcomes were not predetermined, but all subjects could use the feedback to learn a rule, which helped them to maximize the accumulated reward paid out to them after the experiment. Observers were instructed that they could obtain each single reward presented to the observed person, if they had to have learned the rule in the test trials between the learning blocks. Therefore, the feedback stimuli were of high relevance for both active and observational learners. Rather than reduced involvement in the task, the commonality between the study by Yeung et al. (2005) and the present study is the systematic variation of the necessity to adapt the behavior. In both situations—if the subject has no choice between 2 outcomes or if the choice is made by someone else—negative feedback does not signal the necessity of behavioral adaptation, and hence FRN amplitude is reduced. As the present study shows, this does not impair the ability to learn from the feedback an observed person receives, suggesting that—at least in part—other mechanisms are involved in learning by observation.

Evidence from functional imaging further supports the view that a sense of agency affects feedback processing. The finding that near misses, that is, misses proximal to hits, activate the reward system holds only for subjects who actively take part in a gambling task as compared with a passive condition (Clark et al. 2009). Depending on the active engagement of a subject, different striatal subregions are recruited in feedback processing in a learning task. The ventral striatum codes a prediction error irrespective of the behavioral context, whereas the dorsal striatum only comes into play for feedback depending on subjects’ responses (O’Doherty et al. 2004). Thus, it seems reasonable to assume that active and observational learning are mediated by a partly overlapping neural network, with some structures being exclusively or at least more strongly involved in the one type of learning or the other. This notion, however, needs to be further investigated, ideally with patients suffering from selective brain lesions, to provide unequivocal evidence for brain regions particularly recruited during observational learning.

FRN amplitude was also modulated by reward expectancy in the present study. This modulation was driven by the neural response to rewarding rather than nonrewarding outcomes: FRN amplitude was larger for less probable compared with more probable positive outcomes exclusively in the postlearning phase. At the same time, this effect did not interact with group, suggesting that it was not significantly affected by the type of learning. Modulations of reward—rather than nonreward-related ERP amplitudes in the time window of the FRN have been reported in a number of recent studies (Cohen et al. 2007; Eppinger et al. 2008; Holroyd et al. 2008), and Holroyd et al. (2008) have introduced the term correct-related positivity to emphasize that these modulations may in fact be caused by a positive component rather than the FRN.
To summarize, the present study is the first to examine feedback processing in active and observational learning in a between-group design. Clear differences in feedback processing emerged for the active and observation conditions. The FRN following negative monetary feedback (nonreward), which signals the need for behavioral adaptation, was significantly more pronounced in active learners. Similarly, the positive component preceding the FRN was also more pronounced in active learners, with larger amplitude differences between reward and nonreward than in observers. These differences in neural coding did not affect learning. Both active and observational learners gained insight into reward contingencies.

Funding
Ministry of Innovation, Science, Research and Technology of the federal state of Nordrhein-Westfalen, Germany (young researcher programme, Ministerium für Innovation, Wissenschaft, Forschung und Technologie—Programm zur Förderung von Nachwuchsforchergruppen); the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung, grant number 01 GW0541).

Notes
We thank the Ministry of Innovation, Science, Research and Technology of the federal state of Nordrhein-Westfalen, Germany, and the German Ministry of Education and Research for supporting this research.

Conflict of Interest
None declared.

References


Potts GF, Martin LE, Burton P, Montague PR. 2006. When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. J Cogn Neurosci. 18:1112–1119.


