Surprisingly Correct: Unexpectedness of Observed Actions Activates the Medial Prefrontal Cortex

Anne-Marike Schiffer,[1]* Kim H. Krause,[2] and Ricarda I. Schubotz[3]

[1]*Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom*
[2]*Max-Planck-Institute for Neurological Research, Motor Cognition Group, Cologne, Germany*
[3]*Westfälische Wilhelms-Universität,Institut für Psychologie, Arbeitseinheit Biologische Psychologie, Münster, Germany*

Abstract: Not only committing errors, but also observing errors has been shown to activate the dorsal medial prefrontal cortex, particularly BA 8 and adjacent rostral cingulate zone (RCZ). Currently, there is a debate on whether this activity reflects a response to the incorrectness of the committed action or to its unexpectedness. This article reports two studies investigating whether activity in BA 8/RCZ is due to the unexpectedness of observed errors or the incorrectness of the specific observed action. Both studies employed an action observation paradigm reliant on the observation of an actor tying sailing knots. The reported behavioral experiment delivered evidence that the paradigm successfully induced the expectation of incorrect actions as well as the expectation of correct actions. The functional magnetic resonance imaging study revealed that unexpectedly correct as well as unexpectedly incorrect actions activate the BA 8/RCZ. The same result was confirmed for a coordinate in the vicinity that has been previously reported to be activated in separate studies either by the error observation or by the unexpectedness of committed errors, and has been associated with the error-related negativity. The present results suggest that unexpectedness has an impact on the medial prefrontal correlate of *observed* errors. *Hum Brain Mapp 00:000–000, 2013*.

Key words: unexpectedness; mPFC; error observation; action observation; BA 8; fMRI; expectation

INTRODUCTION

How the brain perceives and codes for errors is a long-standing debate of neuroscientific research. As a sideline, some studies have investigated the similarities and differences between errors and observed errors [Bates et al., 2005; Behrens et al., 2008; de Bruijn et al., 2009; Koban et al., 2010; Miltner et al., 2004; Schie et al., 2004; Yu and Zhou, 2006]. None of these paradigms have focused on one particular point: other agents' errors are usually unexpected by the observer.

Some studies have investigated the neural correlates of observed as opposed to committed errors [Bates et al., 2005; Behrens et al., 2008; de Bruijn et al., 2009; Koban et al., 2010; Miltner et al., 2004; van Schie et al., 2004; Yu and Zhou, 2006]. Like many studies on error commission, these studies on error observation typically found activity associated with the medial prefrontal cortex (mPFC) [Behrens et al., 2008; de Bruijn et al., 2009; van Schie et al., 2004]. However, given the recent results that propose that the neural correlates of error commission are rooted in the unexpectedness of errors [Alexander and Brown, 2011; Oliveira et al., 2007; Wessel et al., 2012], one aspect of error observation demands attention: usually, observed errors cannot be fully predicted. This is very important as action observation relies on internal forward models [Flanagan

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Anne-Marike Schiffer, Department of Experimental Psychology, University of Oxford, South Parks Road, OX1 3UD, United Kingdom. E-mail: Anne-Marike.Schiffer@psy.ox.ac.uk

Received for publication 31 October 2012; Revised 28 January 2013; Accepted 3 February 2013

DOI: 10.1002/hbm.22277

Published online in Wiley Online Library (wileyonlinelibrary.com).

and Johansson, 2003; Keysers and Perrett, 2004; Kilner et al., 2004]. Many studies in humans and macaques suggest that internal forward models constantly deliver predictions on the actions that are most likely to be perceived next [Flanagan and Johansson, 2003; Keysers and Perrett, 2004; Kilner et al., 2004; Schubotz et al., 2012; Zacks et al., 2011]. If the observed agent commits an error, this would naturally present an unexpected event to the observer [Schiffer and Schubotz, 2011]. The only instance in which an error would be probabilistically, if not temporally, expected is if the probability of the observed agent committing an error was high; in this case an observer could be surprised at a lack of error (for illustrative purposes, think of a Laurel and Hardy movie). However, in all implemented experimental settings [Bates et al., 2005; Behrens et al., 2008; de Bruijn et al., 2009; Koban et al., 2010; Miltner et al., 2004; Schie et al., 2004; Yu and Zhou, 2006], and in our natural environment, errors are less common than successful actions. This makes (another agent's) errors unexpected and could explain the mPFC correlate of observed errors.

This study was set up to determine whether unexpectedness of observed *actions* would activate the medial BA8/rostral cingulate zone (RCZ). The rationale of the study was that to achieve the dissemination of an unexpectedness effect in action observation, it needs to be separated from the correctness of the observed action. To test this unexpectedness effect, we created probabilistic expectations of correct actions and incorrect actions. Actions that were expected to be performed correctly could either turn out correct (expected) or incorrect (unexpected). Actions that were expected to be performed incorrectly could either turn out correct (unexpected) or incorrect (expected). In a behavioral study, we first established that the paradigm does firmly establish these expectations. For the functional magnetic resonance imaging (fMRI) study, we hypothesized that the unexpectedness of both observed correct and observed incorrect actions causes activity in the medial BA 8/RCZ.

METHODS

Both the behavioral and the fMRI study employed the same stimulus material and very similar paradigms. We will therefore first give a comprehensive introduction to the common setup of both studies (refer also to Fig. 1). Two separate descriptions in greater detail and with a focus on the analyses we employed will then succeed. The main goal of the behavioral study was to test whether it is possible to implicitly create the expectations that either (a) another agent will commit an error or (b) another agent will perform an action correctly. The basic idea of the fMRI study was to investigate the neural responses to violated expectations on whether another agent would perform a correct action or commit an error. The studies were approved by the local ethics committee of the Westfälische Wilhelms-Universität Münster and in line with the Declaration of Helsinki.

Stimulus Material

Both studies employed the videos of an actress tying sailing, climbing, and fishing knots. Using these knots as stimulus material offers a number of advantages that movies of everyday actions do not possess:

1. It is very difficult, if not impossible, for an actor to intentionally perform erroneous actions in a natural way. We trained the actress to the same extent in tying incorrect and correct knots, counterbalancing whether she would learn the correct or incorrect version first. This allowed us to create movies that contained natural looking errors.
2. On a related note, the errors in the knot movies could not be anticipated before they occurred; a danger of intentionally incorrect everyday actions is the possibility that the observer could pick up on differences between a natural action and an intentionally erroneous action before the error occurs. The knot movies, therefore, provide the possibility of timing the onset of the unexpected event.
3. Using colored ropes (four different colors: yellow, green, red, and blue) in the action movies allowed implementing a salient cue during implicit learning whether an action would be performed correctly or not (e.g., red bowline: error; blue cleat hitch: correct). An implementation in real-life actions (e.g., always dropping the red kettle) would have more far-reaching semantic or episodic implications.
4. Real-life errors can be of various severities, reaching from small imperfections in motor trajectories to dropping a Ming vase. The knots allowed to implement a fixed subset of error types: (a) incorrect repetitions of one action step, (b) incorrect skipping of an action step, and (c) performing an action step that is not part of the present knot. They also allowed us to control the error's severity in terms of action outcome. No incorrect knot had a disastrous outcome.
5. In close keeping with the last point, people may differ on what they perceive as an error, and what they would generously call a correct action. Knots can be clearly defined as correct or incorrect. We trained participants to perform knots correctly and controlled their ability to detect errors to achieve an objective measure.

The videos were shot from the first-person perspective, with the camera mounted over the actress' head. The videos showed the actress' lower arms and hands, a board with a cleat fixed to it and the rope the actress was using. Each single knot (16) was filmed 18 times in every color combination (four colors, and white), one time in the correct and one time in the erroneous version. This allowed us to use each single movie only once, even if the condition (e.g., red bowline, incorrect) was shown 10 times over the course of the experiment. This approach was

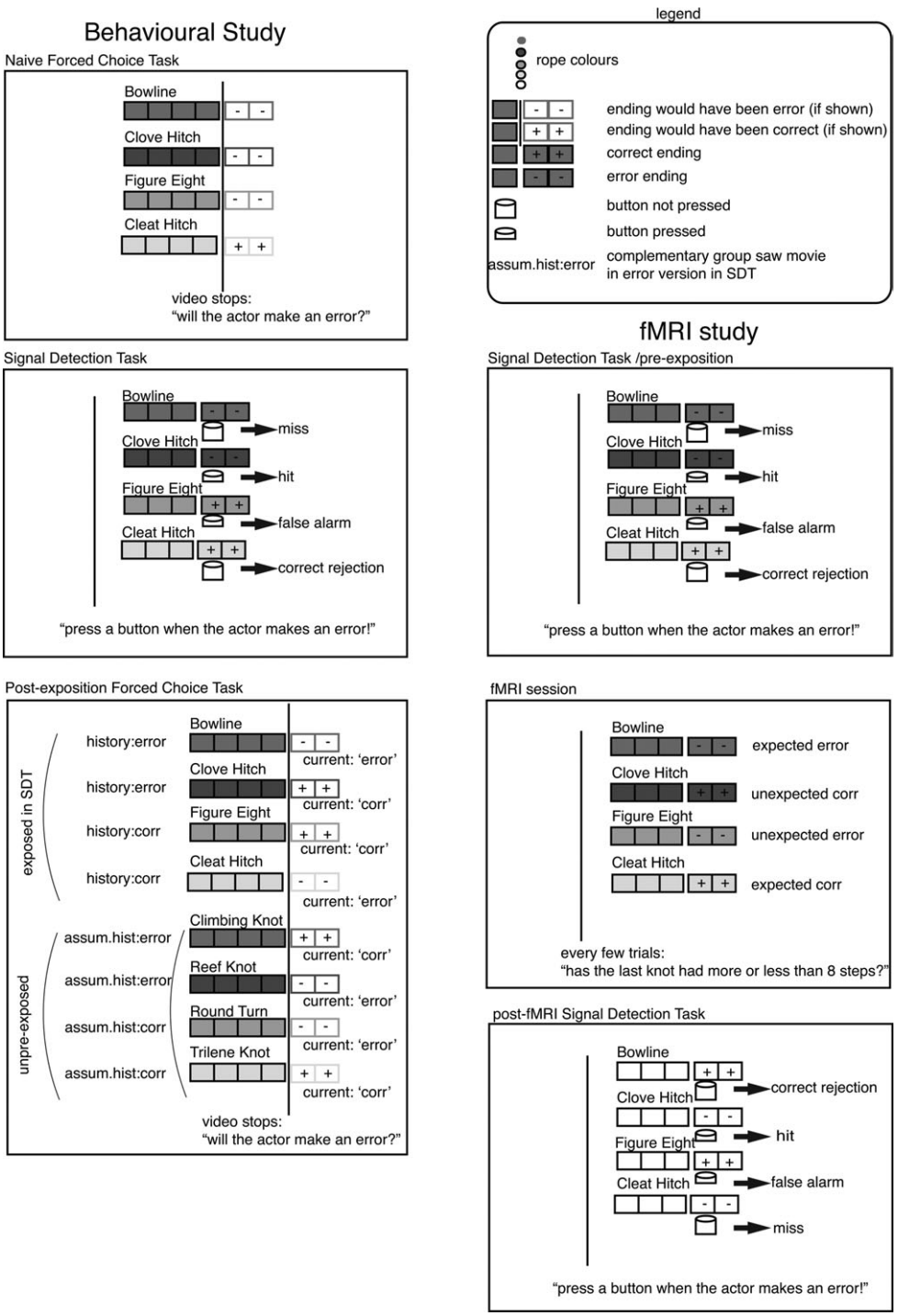

Figure 1.

Comparison of the design of the behavioral study (left column) and fMRI study (right column). The behavioral experiment consisted of the naïve forced-choice task, SDT and postexperimental forced-choice task. The fMRI study included the same type of SDT, followed by the fMRI session that implemented an implicit task (judgment on the number of included action steps).

taken to minimize surface similarities between repetitions that could otherwise account for some of the neural signal measured toward violated expectations. For each movie of each knot, we individually assessed the point in time when it became apparent that the knot would be performed correctly or incorrectly. At this point, it was clearly apparent for 14 out of the 16 knots, which one was presently tied in the video. Two knots had a very similar beginning, but for those knots we took care in the randomization that the color codes would guarantee that participants could determine which knot they presently saw.

Behavioral Training—Behavioral and fMRI Study

Both the behavioral study and the fMRI study involved the same type of training. In a 1-h session, participants were trained to tie 8 out of the 16 knots. Training was based on the photographs of the eight knots. Each knot was presented action step by action step, each one on a separate photograph. The photographs showed tying of the knot with the white rope. The other eight knots of the set of 16 were watched only during training. In these training videos, the knots were tied with a white rope. All videos showed the correct version of the respective knot. Which subset was practiced by tying the knots (motor condition) and which subset was only visually trained (imagery condition) was counterbalanced across participants, with two sets of equally difficult knots. The division between imagery condition and motor condition was implemented because we assumed that most knots that were later shown in their error version had never been practised/acted out in this error version before. This would confound error knots with untied knots. Using an imagery condition allowed deconfounding the influence of no history of tying the knot in the observed fashion (lack of motor model) and the influence of the incorrectness per se. Participants received a CD containing the training videos and the training photographs, written stepwise descriptions of all 16 knots, a board with a cleat mounted onto it, and two white ropes to take home. Participants were instructed to spend overall 10 h between the training session and the experimental session learning the knots. The knots in the motor condition were to be practiced during 5 h of these training hours. The knots in the imagery condition were to be watched for 5 h on video-implementing motor imagery (refer the following sections also). Participants were strictly advised never to tie the knots from the videos. The instruction concerning this visually trained subset was that they should engage in motor imagery. That is, they were told to watch one video, close their eyes, imagine the knot step by step, and then watch the video again to check if they had imagined the knot correctly. The written description of each knot that accompanied the photographs as well as the movies emphasized the number of action steps that each knot should be divided into. Participants were told that they would be tested on their knowledge of the knots and would only receive full compensation if they had perfect knowledge of all of the knots (they did not know that there would be a motor test).

Subjects

Participants either partook in the pilot or in the fMRI session. No participant did both experiments. Out of the 23 healthy participants (17 female, age, 20–48 years, mean [*M*] = 25.39, standard deviation [*SD*] = 7.82), who took part in the pilot study, one was excluded from further analyses owing to insufficient performance (below 2 SDs from mean). Four sets of data were inadvertently deleted after a computer failure and were thus not included. The data from the remaining 18 participants (14 female, age 20–48 years, $M = 25.33$, SD = 8.49) were included in all further analyses. Sixteen healthy participants (11 female, age 20–28 years, $M = 23.25$, SD = 2.46) took part in the fMRI study. Only 15 participants were included in final analyses (11 female, age 20–28 years, $M = 23.4$, $SD = 2.47$), as one participant was excluded from all analyses owing to poor performance during the fMRI. All participants in the behavioral study and 12 out of the 15 participants in the fMRI study worked for course credit. They could earn up to 10 h of course credit for correct performance of all 16 knots. They would get half an hour less course credit for each knot that they could not perform. Participants were paid in course credit for the training sessions in the behavioral study and the fMRI study and received 20 Euro for participation in the fMRI session.

Behavioral Study

The behavioral study session was subdivided into four parts. The first part consisted of an assessment of the subjects' proficiency in tying the 16 knots (motor test). To that end, participants were filmed while tying the knots as fast as they possibly could and later rated by two independent raters according to a predefined set of criterions, for example, correctness of the result, motor fluency, and number of self-corrections. The time-emphasizing instruction was used to heighten error likelihood and achieve a more critical measure of performance proficiency. Only participants who could tie more than eight knots without any errors were allowed to continue in the experiment.

Naïve forced-choice task

The second part of the behavioral session consisted of a forced-choice task on the knot movies (Fig. 1). Two different color versions of each knot were presented. The two versions were either both error versions, both correct

versions, or one correct and one error version The appearance of all knots in the two color version was pseudo-randomly distributed across the length of the experiment. These movies stopped 500 ms before the movements between correct and incorrect versions of one knot diverged. Participants had to press one of the two response buttons to indicate whether they predicted the movie to later contain an error or not ("Will she make an error?"). This task served the purpose of determining whether the actress' intention to commit an error was visible prior to error occurrence. We recorded the standard measures of signal detection theory [Green and Swets, 1974] to estimate *d* prime, an indication of participants' ability to discriminate signal trials (incorrectly tied knots) from other trials (correct knots). We hypothesized that *d* primes would yield no indication of an ability to predict whether the knots would turn out incorrectly or correctly (during the not-displayed ending).

Signal detection task

After the forced-choice task, participants entered the signal detection task (SDT) (Fig. 1). They again watched videos of the knots being tied, but this time, the entire length of the video was shown. Each knot appeared in two different color versions (e.g., red and blue). These two versions were either both error versions, both correct versions, or one correct and one error version. Each version appeared in five different shots, that is, although participants watched, for example, the yellow incorrectly tied bowline five times over the course of the experiment, each single movie was another shot of the same content. For each knot, it was randomly assigned whether it would be shown in the correct or in the error version and what color the rope had. The randomization assured that each rope color would appear equally often with correct or error knots. During this task, participants were required to press a button whenever they detected an incorrectly tied knot. This task was aimed at two principal goals: we used the task as a pre-exposition of knot/color/correctness combinations to ultimately test in the fourth part of the experiment whether participants would learn implicit expectations on the correctness of each color/knot combination. At the same time, the SDT delivered a test for the participants' ability to differentiate between correct and erroneous knots. To measure this, we again used *d* primes. In addition, we measured reaction times over repetitions, to assess learning. Reaction times were taken as the time between the movie-specific point in time when correct knots diverged from error knots and the button press response.

We hypothesized that participants would achieve a high *d* prime, indicating a good ability to detect errors and reject correct movies. We moreover hypothesized that learning would be reflected in decreasing reaction times.

Postexposition forced choice task

Finally, participants underwent a repetition of the first task with an extended set of videos (Fig. 1). Videos were again stopped 500 ms prior to the point of divergence between correct and incorrect knots. In total, 64 different knot/color combinations appeared, with each knot in four different colors. Half of the videos contained knots that had been shown five times before in the SDT, either every time in the correct or every time in the error version. The other half of the videos had not been displayed in the SDT. Participants were again asked to predict and indicate whether the remainder of the video would contain an error or not. This part of the experiment served the purpose of assessing whether the pre-exposition created implicit expectations, leading participants to incorrectly predict videos they had previously witnessed to develop erroneously to contain an error, and vice versa for the previous correct knots. The analysis rested on a comparison of the number of history-driven responses for those movies that had been displayed in the SDT, compared with the number of history-driven responses for the movies that had not. In the latter case, history was defined as the version that the group with the complimentary randomization would have seen during the SDT. The full-factorial model thus contained the two-level factor display in the SDT (displayed or not displayed), the two-level-factor influence of history (response in line with history or different from history), and the two-level factor influence of current version (response in line with the present movie, or different).

We hypothesized that display and history would show a significant interaction in explaining response behavior. We assumed that the pre-exposition of the movie in the SDT (display) would lead to responses in line with the history of the movie.

fMRI Study

Training and admission procedures for the fMRI study were identical to those described in **Behavioral Training—Behavioral and fMRI Study** section. For each imperfect or unfinished knot in the motor test, participants received half an hour course credit or 8 Euro less than the maximum of 10 h/80 Euro. Previous to the fMRI session, the participants underwent the same assessment of their ability to tie the knots as the participants in the pilot study (motor test). The fMRI study also employed an identical SDT as a pre-exposition phase outside the scanner (32 color/knot combinations, each shown five times, responses to observed errors; Fig. 1). Again, the SDT served several purposes: first of all, it allowed us to create implicit expectations on correctness or incorrectness of observed knots (as will be shown in the **RESULTS** section of the behavioral study). Second, we could use the responses in the SDT as a measure of participants' ability to detect errors and reject correct movies as error-free. During the fMRI,

half of the color/knot combinations that appeared repeatedly correctly in the pre-exposition (8) were now displayed in their error version (unexpected error). The other half (8) of the color/knot combinations that appeared repeatedly correctly in the pre-exposition were now again displayed in the correct version (expected correct). Half of the color/knot combinations that always appeared erroneously in the pre-exposition were now displayed erroneously again (8) (expected error), whereas the other half was now displayed in the correct version (8) (unexpected correct; cf. Fig. 1). All movies were shown in full length, no movie stopped before the correctness of the knot was ultimately visible. The randomization that was employed ensured that no color was correlated significantly with one development (i.e., incorrect in pre-exposition and fMRI, correct in pre-exposition and fMRI, switch from correct to incorrect or vice versa between pre-exposition and fMRI). The same logic as for the colors applied to individual knots. The four main conditions (expected correct, expected error, unexpected correct, and unexpected error) contained to equal parts movies from the motor and perceptual training condition. We planned two principal analyses comparing unexpected correct and unexpected error with expected correct and expected error knots. The first was a conjunction of the two unexpectedness contrasts, as will be described in more detail in the following sections. The second analysis aimed to establish whether an ROI, previously reported in the literature, to be activated by error observation was also susceptible to the effects of unexpectedness, dissociated from correctness. This ROI analysis was based on a peak coordinate taken from the study of de Bruijn et al. (2009) errors in action observation study that was mirrored as a peak-activation in the recent study of Wessel et al. (2012). Both our analyses were essentially calculated from the same design matrix and involved the same preprocessing. The analyses were submitted to corrections for multiple comparisons by permutation analysis. We hypothesized that the medial BA 8, potentially reaching into the RCZ [Ridderinkhof et al., 2004] would be activated in the contrast.

fMRI session

During the fMRI session, participants lay supine on the scanner bed. Their head and arms were stabilized using form-fitting cushioning and their hands rested on a rubber foam tablet. On the right-hand side, a response panel was mounted on the tablet and fixed with tape. They could respond to questions with their right-hand index and middle finger resting on two response buttons. The task-irrelevant questions that participants had to answer concerned the number of action steps that the last knot they had seen had encompassed. The last knot could, for example, have been the bowline and the number 8 would appear on the screen. Participants had to press the right button if they judged the knot to have more than eight action steps and the left button if they judged it to have less. Participants wore earplugs to attenuate scanner noise and headphones to listen to the experimenter. Participants saw the display on a mirror built into the head-coil and adjusted individually to allow for comfortable view of the entire screen. The movies extended to 5° of visual angle on the mirror image of the computer screen. Eight null-events of 6 s length were displayed, consisting of the display of a gray background on the screen. Participants were instructed to relax during null-events.

Data acquisition

The functional imaging session took place in a 3T Siemens Magnetom Trio scanner (Siemens, Erlangen, Germany). In a separate session, prior to the functional MRI, high-resolution 3D T-1 weighted whole-brain MDEFT sequences were recorded for every participant (128 slices, field of view, 256 mm; 256 × 256 pixel matrix, thickness, 1 mm; spacing, 0.25 mm). The functional session engaged a single-shot gradient echo-planar imaging sequence sensitive to blood oxygen level-dependent (BOLD) contrast (28 slices, parallel to the bicommisural plane, echo time, 30 ms; flip angle, 90°; repetition time, 2,000 ms; serial recording). Following the functional session immediately, a set of T1-weighted 2D-FLASH images was acquired for each participant (28 slices, field of view, 200 mm; 128 × 128 pixel matrix, thickness, 4 mm; spacing, 0.6 mm; in-plane resolution, 3 mm × 3 mm).

fMRI data analysis

Functional data were offline motion-corrected using the Siemens motion protocol PACE (Siemens, Erlangen, Germany). Further processing was conducted with the LIPSIA 1.6 software package [Lohmann et al., 2001]. Cubic spline interpolation was used to correct for the temporal offset between the slices acquired in one scan. To remove low-frequency signal changes and baseline drifts, a 1/120-Hz filter was applied. The matching parameters (six degrees of freedom, three rotational, three translational) of the T1-weighted 2D-FLASH data onto the individual 3D MDEFT reference set were used to calculate the transformation matrices for linear registration. These matrices were subsequently normalized to a standardized Talairach brain size (x = 135 mm, y = 175 mm, z = 120 mm; Talairach and Tournoux, 1988) by linear scaling. The normalized transformation matrices were then applied to the functional slices, to transform them using trilinear interpolation, and to align them with the 3D reference set in the stereotactic coordinate system. The generated output thus had a spatial resolution of 3 mm × 3 mm × 3 mm. The statistical evaluation was based on a least-square estimation using the general linear model for serially autocorrelated observations [Worsley and Friston, 1995]. Temporal Gaussian smoothing (4-s FWHM) was applied to deal with temporal autocorrelation and determine the degrees of freedom [Worsley and Friston, 1995]. Spatial

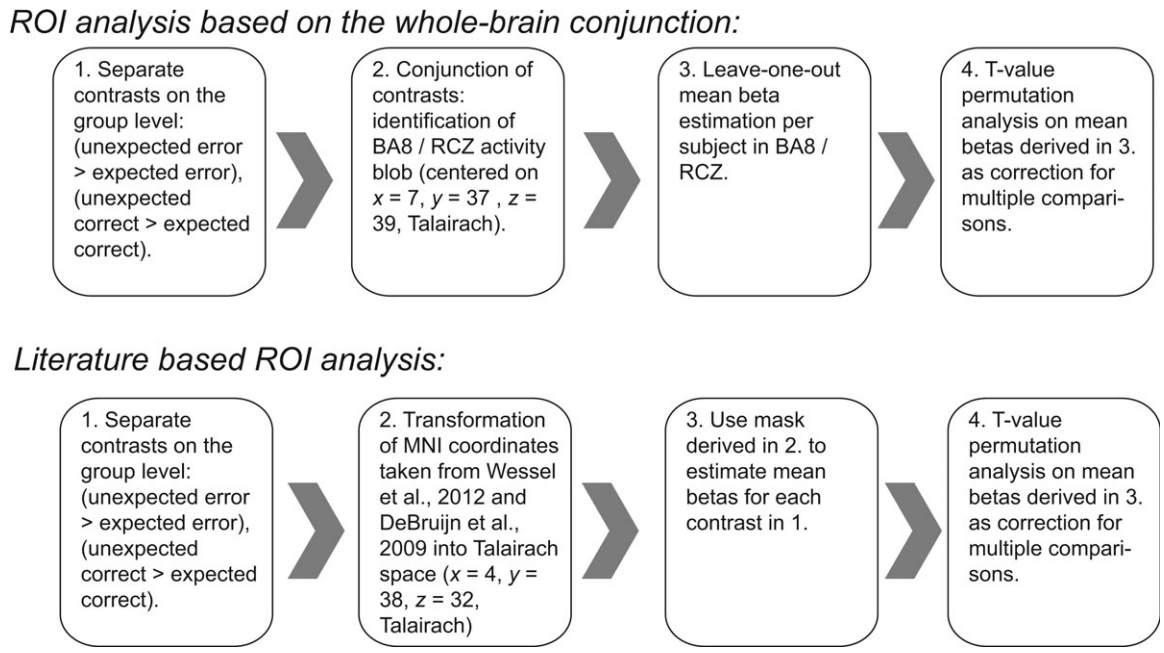

Figure 2.
Overview of the analysis steps for the ROI analysis based on the whole-brain conjunction and the literature-based ROI analysis.

smoothing of 5.6 mm FWHM was applied. The design matrix was generated by hemodynamic modeling using a δ-function (stick function). The design matrix encompassed the following events: unexpected correct knot movies, unexpected error knot movies, expected correct knot movies, expected error knot movies, and null-events. The onset that was modeled in the design matrix was exactly the movie or null-event onset. All events were modeled with an amplitude vector of 1. The entire movie length was modeled for each knot movie and the null-events were modeled by their actual length of 6 s.

Whole-brain analysis (conjunction analysis)

To test for the shared effects of unexpectedness, we employed a conjunction analysis. Conjunction analyses deliver a possibility to ensure that both comparisons contribute to the final activity pattern. The analysis consisted of the overlay of the contrasts (unexpected correct > expected correct) ∩ (unexpected error > expected error). The conjunction of contrasts was performed at a probability level of $P = 0.05$ threshold and only activations higher than $z = 2.0$ are reported. As a follow-up test, we used an ROI analysis that employed a jack-knife correction to prevent double-dipping [Egner, 2011; Kriegeskorte et al., 2009] and a permutation analysis for second-level corrections (Fig. 2, Nichols and Holmes, 2002; Nichols et al., 2005). The jack-knife approach rests on an iterative leave-one-out technique (Fig. 3). This technique estimates the peak voxel in a contrast n times (n, number of subjects), each time with one of the subjects excluded from the analysis. The peak voxel that is estimated without the specific subject is then applied as a ROI to the subject that was left out [Egner, 2011; Nee and Brown, in press]. We used the peak voxel from the (leave-one-out) conjunction analyses to estimate the mean β for each subject (the respective left-out-subject) in both original contrasts that made up the conjunction. The two samples of mean βs for the population of subjects were then fed into a permutation analysis [Nichols and Holmes, 2002; Schiffer et al., 2012]. The permutation analysis employed 30,000 random assignments of mean βs to estimate the true distribution of betas in the ROI. This bootstrapping approach does not necessitate the assumption of a Gaussian distribution within an ROI and it does not artificially boost the power of activity in small ROIs [Nichols and Holmes, 2002]. It is, hence, more conservative than other second-level corrections for multiple comparisons. We expected an increased BOLD signal in the RCZ, bordering the medial BA 8 for unexpectedness, deconfounded from the correctness or incorrectness of the action. We expected that we would be able to substantiate this result in both contrasts that entered the conjunction analysis in the combined jack-knife permutation analysis approach, delivering a significant result, corrected for

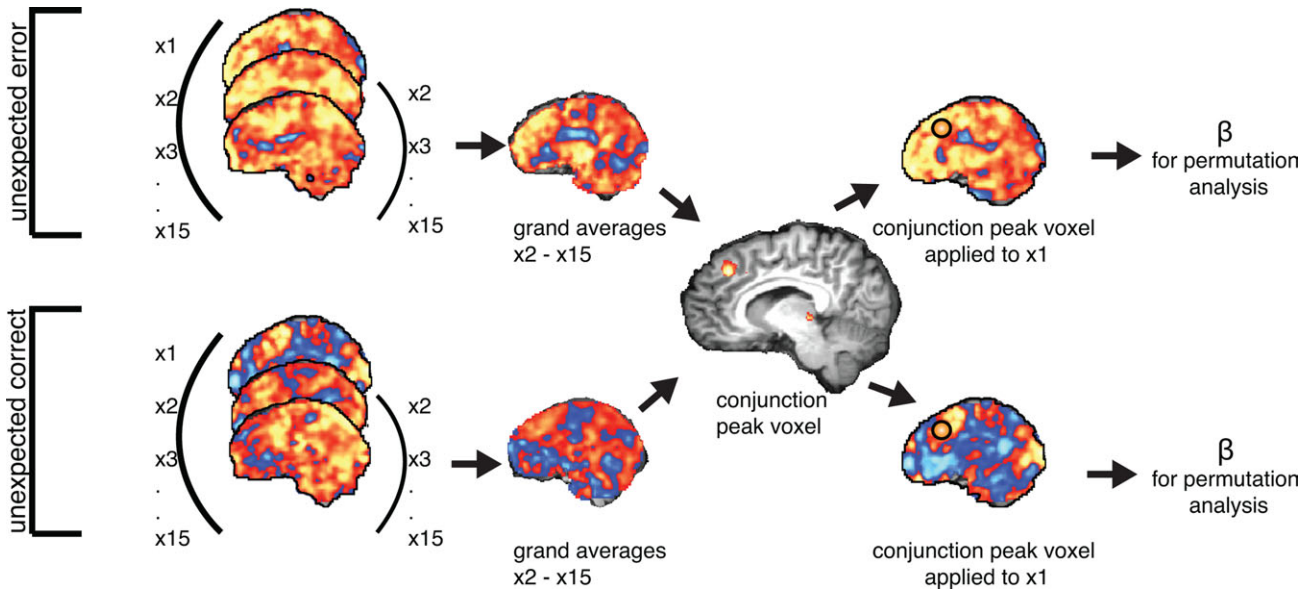

Figure 3.

Leave-one-out approach: The overlap of activity on the group level of a subset of the entire group leaving out one brain is used to extract the medial prefrontal ROI. The peak coordinates are subsequently used to extract the mean β-value within both contrasts of the conjunction in the left-out brain. This process is repeated for every brain to be left out—and the mean β thereof extracted—once. Resulting βs are used in the permutation analysis.

multiple comparisons both for unexpected correct and for unexpected error knots.

Literature-based ROI

The second ROI analysis was planned in case that the mPFC peak voxel of the conjunction analysis would not coincide with the peak voxel reported previously for observed errors. The motivation for this further analysis was twofold: first of all, we wanted to test whether our data are consistent with the recent literature of error processing [Wessel et al., 2012] that emphasizes the unexpectedness account. Furthermore, we wanted to test whether the explanation of unexpectedness provides an alternative to the explanation of incorrectness for coordinates reported for the processing of other agents' errors [de Bruijn et al., 2009]. De Bruijn and colleagues reported that errors, either of the individual itself or observed by the individual, lead to an activation around the peak coordinate at $x = 4$, $y = 33$, $z = 42$ in MNI space. Wessel and colleagues reported a similar (albeit left lateralized) peak coordinate for unexpected events (errors and novel items) at $x = -6$, $y = 33$, $z = 42$ in MNI space. For the purpose of the ROI analysis, we transposed the MNI coordinates to Talairach space, using the program mni2tal.m, run on Matlab 2010a (The Mathworks). The resulting Talairach coordinates were $x = 4$, $y = 28$, $z = 38$ (corresponding to de Bruijn et al., 2009) and $x = -4$, $y = 28$, $z = 32$ (corresponding to Wessel et al., 2012). Both coordinates and three surrounding voxels (i.e., a 27 mm^3 ROI centered on the peak coordinate) entered the analysis as one ROI. We calculated the mean β-values for every participant in the ROI in both contrasts: (unexpected correct > expected correct) and (unexpected error > expected error). These contrasts are parallel to the contrasts that entered the conjunction analysis in the whole-brain conjunction-derived ROI. The aim of this analysis is thus to test for the significance of both unexpectedness effects in the ROI derived from the literature. Literature-derived ROIs do not call for a leave-one-out approach [Kriegeskorte et al., 2009]. All later steps of the analysis were conducted in parallel to the conjunction-derived ROI described previously (Fig. 2). The permutation analysis of *t*-values in the conjunction analysis-based ROI tests whether each entering contrast yields significant activity in this area. In the same vein, the permutation analysis based on the literature-derived ROI tests whether the same two contrasts each yield significant activity in the ROI taken from the literature.

In a second approach, we calculated the mean βs for every participant in the contrast (unexpected error ∩ expected error) > (unexpected correct ∩ expected correct) as an additional test for a singular influence of the observations of errors. Finally, to test whether the influence of unexpectedness would outweigh the influence of incorrectness, we calculated the interaction contrast of the two main effects (unexpected error > unexpected correct) >

(expected error < expected correct). For each contrast, the distribution of βs was then fed into a permutation analysis of 30,000 permutations as a corrective for multiple comparisons. The respective *t*-values for both comparisons were then compared against the critical *t*-value ($P = 0.05$; position 28,500 in the distribution) derived from the permutation analysis.

We hypothesized that the ROI would be activated by the unexpectedness of an event. We did not expect a significant activation for the observation of errors. We hypothesized that the effect of unexpectedness would surpass the effect of incorrectness in the interaction analysis.

Explorative observed error search

In addition to the previously reported hypothesis-driven analysis, we calculated the main effect of error (unexpected error ∩ expected error) > (unexpected correct ∩ expected correct) on the whole-brain level. This whole-brain analysis was calculated for two reasons: first, as the task in the fMRI did not yield a measure of the participants' awareness of errors, we rely on a neuronal measure of a difference between erroneous and correct actions. Finding this difference would substantiate the study's claim to yield the results that could potentially relate to the previous findings in error-observation studies. Second, masking this contrast with the conjunction for unexpectedness at a threshold of $z = 1.96$ allows to probe whether the observation of errors activates the mPFC to a degree that surpasses the activity in the conjunction of unexpectedness. Arguing in the same vein, we also calculated the error contrast only for unexpected movies (unexpected error > unexpected correct). This contrast can be understood to capture the element of error, freed from the element of unexpectedness. A significant finding in this contrast in the described ROI ($x = 4$, $y = 28$, $z = 38$) would indicate that, at least for unexpected movies, incorrectness itself creates a difference in the ROI. This potential finding would again implicate incorrectness as a modulatory influence on unexpectedness area. Mean βs were taken for each participant in the ROI and the critical *t*-value was determined by means of the previously described permutation analysis.

fMRI-Post-test

The fMRI-session was followed by a brief post-test. Participants sat in front of a laptop and saw each of the 16 knots again tied once with the white rope (i.e., the training color), either in the correct or in the erroneous version. Participants had to indicate whether the knot they saw was tied correctly or incorrectly. This test was meant to assess whether participants were still able to discern correct from incorrect knots after the pre-exposition and fMRI. This measure is important to ensure that participants did not unlearn the knots. Only if unlearning could be ruled out, potential error and unexpectedness effects measured during the fMRI would remain meaningful.

TABLE I. Mean RTs in seconds (s), SD of RTs (s), and mean *d* primes for all behaviorally assessed parts of the behavioral and fMRI study

	RT—mean (s)	RT—SD (s)	*d* Prime
Behavioural study			
Naïve forced-choice	0.93	0.18	0.09
SDT	2.42	0.57	2.33
Postexposition forced-choice	0.75	0.22	−0.05
FMRI			
SDT (pre-exposition)	2.72	0.51	3.86
FMRI—questions	0.93	0.29	

RESULTS

Behavioral Study

The behavioral study was set up to assess whether we can implicitly train subjects to expect whether another actor would tie knots incorrectly or correctly. In total, 18 participants met the initial motor test criterion. They performed on average 11 knots correctly (rounded figure; $SD = 3.07$). These participants were then allowed to participate in the study.

Naïve forced-choice task

The naïve forced-choice task was implemented as a test of the stimulus material. If participants had displayed the ability to discriminate between incorrect and correct knots before the error occurred in the respective movies, this would have implied that error and correct knots were too dissimilar and that the time point of error commission could not be assessed precisely. However, participants' *d* primes in this task showed no evidence of the ability to discriminate between target (error) and distractor (correct) knots (*d* primes and RT, Table I). A one-sample *t*-test against zero yielded a *t*-value of 0.2 and $P > 0.05$. Thus, we got no indication of the possibility to tell these movie types apart before the error was committed.

Signal detection task

The SDT was used to assess participants' ability to tell error and correct knots apart. Its implicit function was to train specific expectations concerning the correctness or incorrectness of specific color/knot combinations. Participants' *d* primes in this task suggested a good ability to detect error knots (*d* primes and RT, Table I). A *t*-test of the *d* prime distribution against zero yielded a *t*-value of 10.12 and *P*-value of <0.05. Hence, participants were well able to correctly identify knots either as correct or as incorrect knots.

We calculated reaction times as another measure of learning. Reaction times for error detection for each specific movie decreased significantly over the course of the five repetitions (Table I). A repeated measures ANOVA with repetition as a five-level within subject factor yielded a significant result $F_{(4,68)} = 11.015$ ($P < 0.05$).

Postexposition forced-choice task

The postexposition forced-choice task was meant to estimate whether seeing a knot in the correct or error version during the pre-exposition would create a persistent expectation that this knot would be tied in the error or correct version (d primes and RT, Table I). A t-test of the d prime distribution against zero yielded a t-value of -0.61 and $P > 0.05$. The test for the establishment of expectations was a repeated measures ANOVA with the three two-level factors display (seen before, not seen before), history (in line with response, not in line with response), and current version (in line with response, not in line with response). This repeated measures ANOVA yielded a significant main effect of display with $F_{(1,17)} = 9.983$, a significant interaction between the factor history and the covariate individual d prime with $F_{(1,17)} = 4.564$ in the SDT and an interaction between the factors display, history and the covariate d prime in SDT with $F_{(1,17)} = 16.869$, (all $P < 0.05$). We thus find that the previously seen version influenced the participants' response behavior. More specifically, this influence was scaled by their responses during the pre-exposition. Quite interestingly, participants who had a high d prime during the SDT were less influenced by the history of movies (which they had actually seen). Note that the history of movies that had actually not been displayed to the subject before (half of the movies) was set as the history that the movie had in the other randomization. Including history and display as factors ensured that history effects were not owing to the fact that the actress had at another stage performed the same knot in another condition (possibly resulting in idiosyncratic movements for each knot, or likewise), but that the subject has actually been exposed to that condition. We thus find that if participants had seen a movie in a certain condition (correct or incorrect), this biased them toward expecting the movie to develop in the same way (correct or incorrect, respectively).

fMRI Study

Motor test, pre-exposition SDT, and post-MRI SDT test

In the initial motor test, participants displayed a high ability of tying the knots, tying on average 13 knots correctly (rounded figure; $SD = 2.53$). During the SDT that served as a pre-exposition to the fMRI participants' d primes varied between 1.42 and 6.05 (d primes and RT, Table I). Reaction times analyzed in a repeated measures ANOVA again yielded a significant decrease ($F_{(4,52)} = 10.021$) of RTs ($P < 0.05$). In the post-test that probed participants' uncompromised ability to tell error knots from correct knots, participants' d primes varied between 1.34 and 4.65 (d primes and RT, Table I). A two-tailed paired t-test between the d prime distributions of pre- and post-test did not yield a significant result ($t = 1.05$; $P > 0.05$). Thus, we find no indication that participants' ability to discern correct and incorrect knots changed.

Effect of unexpectedness: Conjunction analysis and jack-knife permutation analysis approach

The conjunction ([unexpected correct > expected correct] ∩ [unexpected error > expected error]) yielded above-threshold activity in the mPFC (medial Brodman Area [BA] 8). This area can be understood as a transitional zone to the RCZ (Talairach coordinates: $x = 7$, $y = 37$, $z = 39$; $z = 2.48$, uncorrected). Corrections for multiple comparisons on the whole-brain level yielded no significant activations (a list of activations of this contrast uncorrected for multiple comparisons is summarized in Supporting Information Table 1).

To test the statistical significance of the recorded activity in the medial BA 8/RCZ, we implemented a jack-knife approach as described in the **METHODS** section (Fig. 3). This approach circumvents a double-dipping fallacy [Egner, 2011; Kriegeskorte et al., 2009]. This approach was followed by a correction for multiple comparisons by permutation analysis. This analysis yielded significant effects for both contrasts, with the (unexpected error > expected error) - contrast yielding a t-value of 3.46, compared against a critical t-value of 1.76 (position 28,500, $P = 0.05$). The (unexpected correct > expected correct) - contrast yielded a t-value of 2.56, compared against a critical t-value of 1.75 (position 28,500, $P = 0.05$) (Fig. 4). Thus, we established a significant effect of unexpectedness in the same ROI in the RCZ both for error and for correct actions that survived correction for multiple comparisons [Nichols and Holmes, 2002].

Literature-based ROI analysis

The second analysis calculated the mean βs for every participant in the literature-based ROI ($x = 4$, $y = 28$, $z = 38/x = -4$, $y = 28$, $z = 38$, Talairach coordinates) for the contrasts (unexpected correct > expected correct) and (unexpected error > expected error). Both contrasts yielded a significant result. The (unexpected error > expected error) contrast reached a t-value of 2.26, compared against a critical t-value of 1.74 (position 28,500, $P = 0.05$). The (unexpected correct > expected correct) contrast reached a t-value of 1.93 compared against a critical t-value of 1.74 (Fig. 4). The main effect of the error - contrast (unexpected error ∩ expected error) > (unexpected correct ∩ expected correct) yielded no evidence for a significant BOLD increase for the observation of errors, irrespective their unexpectedness. The critical t-value in

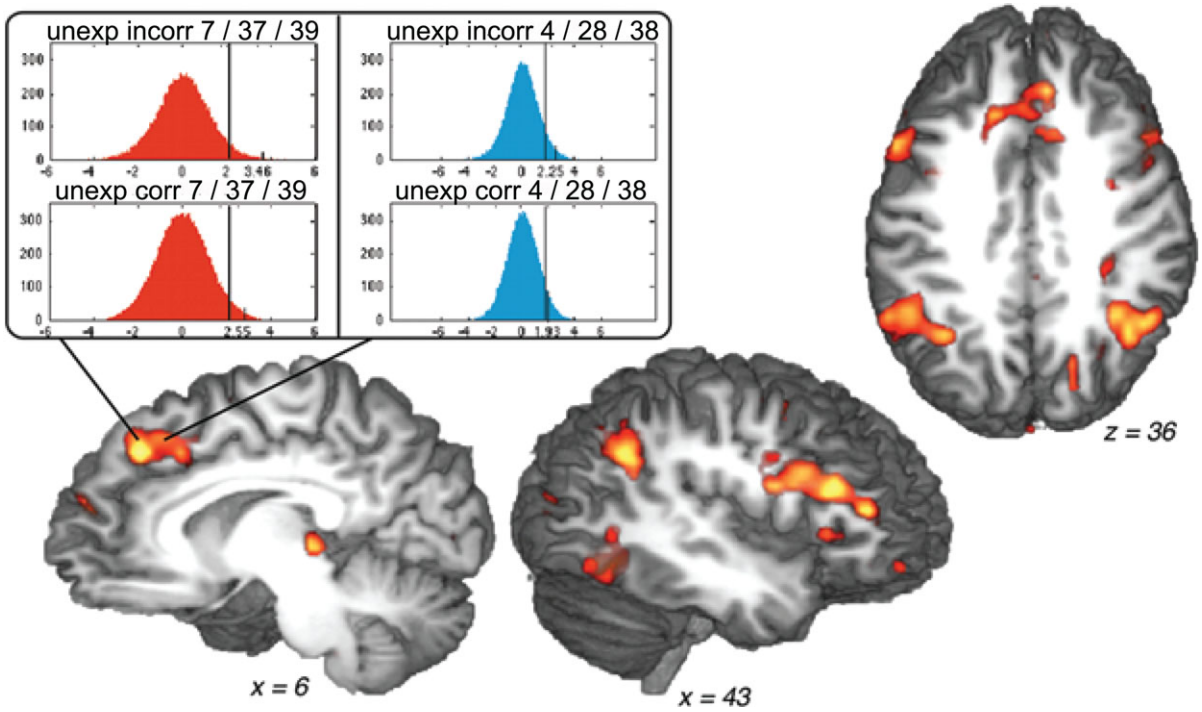

Figure 4.

Whole-brain activations for the conjunction analysis: (unexpected correct > expected correct) ∩ (unexpected error > expected error). The red distributions refer to the *t*-values calculated from the 30,000 permutations in the jack-knife-derived ROIs. The respective critical *t*-value is marked by the long bar that separates the distributions, the exact *t*-value is marked by the small bar (in all cases to the right of the latter), and spelled out on the *x*-coordinate. The blue distributions complementarily show the results for the literature-based ROI permutation analysis. The distributions for unexpectedly incorrect knots are in the top row and the distributions for the unexpectedly correct knots are in the bottom row.

this distribution was 1.72, whereas the main effect of error reached a *t*-value of 0.23 in the literature-based ROI. However, not finding a main effect of error, although the null effect was expected, does not rule out that error quality influences the effect of unexpectedness. Investigating whether errors in fact do not influence the unexpectedness effect is hence better achieved by means of an interaction contrast [Nieuwenhuis et al., 2011]. Therefore, we calculated the interaction between the two putative main effects, the above-mentioned main effect of error and the main effect of unexpectedness, and thus controlling for the error effect: ([unexpected error > unexpected correct] > [expected error < expected correct]).

This interaction analysis yielded a *t*-value of 1.72, and a critical *t*-value of 1.76 (position 28,500, $P = 0.05$). The actual *t*-value of 1.72 took the position 28,463 in the distribution, which corresponds to a *P*-value of 0.0512. Although this does rule out an influence of the incorrectness of the observed action on the unexpectedness effects at $P = 0.05$, we regard the potential error influence still as rather unlikely. It is important to note that the test for the reverse interaction (unexpectedness effects larger for error than for correct actions) would not be statistically sound as it would result in testing for evidence of the null-hypothesis.

Explorative observed error search

The whole-brain analysis for the main effect of error yielded activity in the inferior frontal junction, the posterior intraparietal sulcus, the posterior middle temporal gyrus, and the cerebellum (Table II) at a very lenient correction for multiple comparisons of $Z = 1.96$. Activation in the insula, which may be of particular interest in error-related contrasts, was present in the uncorrected grand averages ($x = -29$, $y = 24$, $z = 12$, Talairach coordinates; $z = 3.007$), but did not survive correction for multiple comparisons. Masking the main effect of error contrast with the conjunction for unexpectedness in errors and correct actions yielded no activity in the mPFC. Finally, the effect of incorrectness only for unexpected actions did not significantly activate the literature-derived ROI ($x = 4$, $y = 28$, $z = 38$; Talairach coordinates). The mean β distribution derived in this ROI for that contrast (unexpected error > unexpected correct) yielded a *t*-value of 0.47, with the according critical *t*-value being 1.80. On the whole-brain

TABLE II. List of highest peaks of activation for the main effect of error (unexpected error ∩ expected error) > (unexpected correct ∩ expected correct) including anatomical location, Talairach coordinates of the peak voxels, and z-values

	Local maxima (Talairach coordinates)			
Area	*x*	*y*	*z*	*z*-Values
Inferior frontal junction (r)	34	7	30	3.37
Posterior intraparietal sulcus (r)	25	−56	30	3.53
Posterior middle temporal gyrus (r)	49	−56	6	3.72
Cerebellum (l)	−29	−62	−39	3.63

Corrected for multiple comparisons at $z = 1.96$.

level, this contrast dealing with the effect of incorrectness in unexpected events activate the right inferior frontal gyrus, right inferior temporal junction, right ventral premotor cortex, right temporoparietal junction, right fusiform gyrus, right parahippocampal gyrus, and the most anterior superior temporal gyrus (all activations corrected for multiple comparisons at $z = 1.95$, $P = 0.05$).

DISCUSSION

The present fMRI study investigated whether the mPFC, particularly BA 8/RCZ, codes for the unexpectedness of observed actions, regardless of whether they are correct or not. In an accompanying behavioral study, we first ensured that implicit learning can create the expectation of correct, as well as the expectation of incorrect actions. In the subsequent fMRI study, we found BA 8/RCZ to code for the unexpectedness of an observed action, regardless of whether this action was a correct one or not. Our findings suggest that activity previously reported in relationship to the observation of errors [Bates et al., 2005; Behrens et al., 2008; de Bruijn et al., 2009; Koban et al., 2010; Miltner et al., 2004; Schie et al., 2004; Yu and Zhou, 2006] could potentially have been driven, to some extent, by their unexpectedness.

This interpretation substantiates recent work that likewise implicated activity in RCZ as driven by the unexpectedness of events [Alexander and Brown, 2011; Holroyd et al., 2009; Wessel et al., 2012]. Regarding error observation, however, most studies measured an event-related potential (ERP) often associated with error commission: the so-called error-related negativity (ERN). It is naturally beyond the scope of this study to prove that the ERN recorded toward observed errors in the previous studies was owing to the unexpectedness of actions and unrelated to the fact that they were erroneous. However, we may suggest that unexpectedness as a characteristic of observed errors likely contributed to the respective ERNs. We will first investigate in how far the previous studies controlled for the effects of unexpectedness, turn toward the ERN as an ERP measure of error (observation), and finally discuss the current findings with respect to the notion of unexpectedness.

Previous Studies on Error Observation

Most studies on error observation use ERPs. These studies target the ERN, a component originally associated with error commission [Falkenstein et al., 1991; Gehring et al., 1993]. Thus, the error-observation ERN has been claimed to reflect an internal error mirror system that responds to observed errors and thus allows observational learning [Bates et al., 2005; Koban et al., 2010; Miltner et al., 2004; Schie et al., 2004]. Quite interestingly, neither fMRI study [Behrens et al., 2008; de Bruijn et al., 2009] implemented the observation of another actor, but the observation of "their" error feedback; the participants learned through feedback on a computer screen whether the virtual agent had made an error or performed the task correctly. This rather indirect setup weakens the argument made by other authors that the observed frontomedian activity could be a correlate of an error-focused extension of an internal error mirror neuron system [Bates et al., 2005; Koban et al., 2010; Miltner et al., 2004; Yu and Zhou, 2006]. The highly debated mirror neurons are supposed to be activated by the perception of other people's actions [Rizzolatti and Sinigaglia, 2010; but cf. Csibra, 2007; Keysers and Perrett, 2004; Kilner et al., 2004] and not by a corresponding abstract action feedback. We take this finding of outcome-related "mirror" activity as a result which the error-mirror neuron account cannot accommodate. On a related note, the bilateral PM activity for unexpected events deserves mentioning. The PM activity in the unexpectedness contrast can be interpreted as a correlate of the mismatch signal between an internal forward model of the observed action and the bottom-up visual input. This mismatch interpretation has been tested and reported, to some extent, in previous studies [Schiffer et al., 2012; Schubotz, 2007; Schubotz and von Cramon, 2004].

Regarding the validity of the proposed unexpectedness account, some previous studies on error observation did control for unexpectedness [Bates et al., 2005; Miltner et al., 2004]. However, in one of these studies, unexpectedness was implemented as immediate trial history [Bates et al., 2005]. The study hence controlled for the possibility that an ERN would be owing to the fact that the last observed error/ERN had occurred some time ago, as opposed to having occurred recently. However, it is not quite clear whether a string of correct trials should lower or heighten the expectation of an error trial (gambler's fallacy). Moreover, it appears as if each series length of correct trials between two incorrect (ERN eliciting) trials occurred equally often [Bates et al., 2005]. This balanced

trial history leads to an equal probability (and expectation) of error trials on each correct trial. The argument of unexpectedness by trial history is thus somewhat weakened. In a similar vein, another study on error observation investigated whether the ERN component would be superimposed on a P300 component for unexpectedness [Miltner et al., 2004]. The fact that error trials did not elicit a P300 was interpreted as an indication that the ERN could not be rooted in the unexpectedness of errors [Miltner et al., 2004]. This argument, however, rests on the assumption that the P300 signals unexpectedness and the ERN does not, an account that has been challenged [Croft et al., 2003]. Interestingly, the reverse pattern—a P300 toward observed errors but no ERN - has been reported [de Bruijn, et al., 2007]. This study on the observation of erroneous actions did in fact establish a P300 correlate toward unexpected observed errors, but no ERN. The authors interpret their P300 finding as an indication of the unexpectedness of errors, challenging the previously reported results.

Localizing Observed Error Responses

If fMRI results are to inform ERP measures and vice versa, it is desirable to map an ERP to a corresponding anatomically specified BOLD increase. An elegant approach that used concurrent fMRI and ERP measures did locate a source for the ERN in the RCZ [Debener et al., 2005] at $x = 0$, $y = 17$, $z = 42$, Talairach coordinates. The considerable Euclidian difference to our peak voxel ($x = 7$, $y = 37$, $z = 39$) renders analogies inconclusive. Another valid approach to allow inferences between ERPs and anatomy is to use the same paradigm in an fMRI and ERP study and to compare the results as Wessel and colleagues did. Their task allowed comparing the BOLD response and ERP components (ERN and N2) toward unexpected events, such as errors and novel events (oddballs). The ensuing comparison established that the BOLD activity in the RCZ, the ERN, and the N2 was closely related and could all be explained substantially by the unexpectedness of an event. To recuperate, the analysis we conducted in an ROI which previously featured in error-observation research [de Bruijn et al., 2009] led us to conclude that even the unexpectedness of correct actions activates this ROI (as well as the unexpectedness of errors) and suggests that the correctness of an observed action does not interact with the unexpectedness signal in this ROI. Future research needs to determine whether the same holds true for the associated ERN.

Unexpectedness and Model Adaptation

The identification of the RCZ as an area activated by unexpectedness gains substantial support from the literature [Alexander and Brown, 2011; Beckmann et al., 2009]. Although our work does not dissociate other aspects of committed errors, it does contribute evidence for unexpectedness effects in the specific area. This unexpectedness signal has been related not only to unexpected outcomes, but also to the omission of an expected outcome [Alexander and Brown, 2011]. This account is close to the concept of an unexpectedly omitted error, as in this study (in unexpected correct actions). Functionally, the unexpectedness signal was suggested to prompt the adjustment of an internal forward model (e.g., of an action) [Holroyd et al., 2009; Ullsperger et al., 2004]. A similar interpretation was put forward in a number of articles concerning the involvement of the medial BA 8 in decisions under uncertainty [Volz et al., 2003, 2004, 2005]. Subjective uncertainty can be regarded as the result of an objective accumulation of unexpected events [Friston, 2010; Luce, 2003]. Thus, the results of this study dovetail with an account of adaptation of internal forward models driven by an unexpectedness signal in the dorsal RCZ. Moreover, this study indicates that the internal forward model that is adapted is not necessarily that of the action in the sense of a motor model, but can relate to an anticipated perception. Participants in both our studies learnt to expect that a specific knot would be tied incorrectly but they did not unlearn the tying of the knot. This result extends the view proposed previously that correlates of the observed errors influence observational learning (e.g., Schie et al., 2004). These accounts focused on the adaptation of the observed action, not on the adaption of the expectation of the performance of the observed action.

Adaptation of internal forward models is closely associated with a type of error that has so far not received due to attention in this discussion: the prediction error. Prediction errors are defined as mismatches between anticipated and actual sensory input (for reviews, see Bubic et al., 2010; Clark, 2012; Friston, 2010). The occurrence of prediction errors incites the adaptation of internal forward models. Importantly, unexpectedness, or surprise [Friston, 2010] captures the same phenomenon: a sensory input not in line with the brain's current predictions. To recuperate: Based on the presented data, we have argued the case that activity in the mPFC toward observed errors can be driven by these observed errors' unexpectedness. The fact that unexpectedly correct movies activate the same area shows that unexpectedness alone can cause this activation. Bringing together these accounts of prediction errors in perception and unexpectedness, we suggest that unexpected events represent errors in the internal forward model. Very speculatively, the RCZ or mPFC may thus be involved in coding for the necessity of adaptation when internal models fail. Be that internal forward models of perception, or even action, when the agent's own errors are involved [Friston, 2010]. If the contrast of errors within unexpected movies had led to a significant activation in the RCZ, this would have challenged this interpretation, as the observation of unexpected errors should NOT lead to more adaptation than observation of unexpected correct actions. However, this contradicting evidence was not obtained.

A very interesting finding regarding the adaptation of an internal forward model is the possible activation of the

habenula (Supporting Information Table 1), a part of the epithalamus. The habenula has often been described as signaling for punishment, or worse-than-expected outcomes [Hikosaka et al., 2008; Matsumoto and Hikosaka, 2007, 2009]. However, we could repeatedly establish habenula activation toward deviations from an internal forward model that were free of intrinsic valence [Schiffer and Schubotz, 2011; Schiffer et al., 2012]. Moreover, both observed errors [de Bruijn et al., 2009] and uncertainty [Volz et al., 2003] may entail activation of the habenula. Especially, the putative activation in the observed errors study is interesting, as the observed errors encompassed errors of the other agent which the observer sometimes benefitted from. Thus, they had an (observed) error quality and possibly an unexpectedness component—but penalty effects were balanced. Importantly, the habenula has the capacity of modulating midbrain dopaminergic output [Hikosaka et al., 2008; Matsumoto and Hikosaka, 2007] and dopamine in turn has been related to the guidance by versus switching away from an internal model in perception [Friston, 2012]. Thus, the deviation signal in the habenula may be involved in the dopaminergic modulation necessary to adjust internal forward models.

CONCLUSIONS

The dorsal RCZ, bordering BA 8, is associated with the unexpectedness of observed events, be they errors or not. This finding extends to an ROI previously implicated separately in coding for unexpected events, observation of errors, and the generation of the ERN. Hence, this unexpectedness account seems more parsimonious than a previous error-mirror-neuron account. It deserves further experimental scrutiny whether unexpectedness, as a sign of a failing internal forward model, draws on the same resources as error coding. Not last because a recent study has discovered separate neuronal populations in the monkey SMA: one seemingly coding for unexpected erroneous actions and one for unexpected outcomes [Yoshida et al., 2012].

ACKNOWLEDGMENTS

The authors express their sincere gratitude to Hilde Haider for her pivotal support of the project. The authors further thank Luisa Donner and Janji Yokeeswaran for their contributions to the design of the experiment, creation of stimulus material, and data collection. The authors thank Christiane Ahlheim for advice on statistical analyses. Finally, the authors also thank Thomas Maraun, who inspired the project in its early days.

REFERENCES

Alexander WH, Brown JW (2011): Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci 14:1338–1344.

Bates AT, Patel TP, Liddle PF (2005): External behavior monitoring mirrors internal behavior monitoring. J Psychophysiol 19:281–288.

Beckmann M, Johansen-Berg H, Rushworth MFS (2009): Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. J Neurosci 29:1175–1190.

Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS (2008): Associative learning of social value. Nature 456:245–249.

de Bruijn ERA, de Lange FP, von Cramon DY, Ullsperger M. 2009. When errors are rewarding. J Neurosci 29:12183–12186.

Bubic A, Von Cramon DY, Schubotz RI. (2010). Prediction, cognition and the brain. Frontiers in human neuroscience, 4.

de Bruijn ER, Schubotz RI, Ullsperger M. (2007). An event-related potential study on the observation of erroneous everyday actions. Cognitive, Affective, & Behavioral Neuroscience, 7(4):278-285.

Clark A (2012): Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci.

Croft RJ, Gonsalvez CJ, Gabriel C, Barry RJ (2003): Target-to-target interval versus probability effects on p300 in one- and two-tone tasks. Psychophysiology 40:322–328.

Csibra G (2007): Action mirroring and action interpretation: An alternative account. In: Haggard P, Rosetti Y, Kawato M, editors. Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII. Oxford: Oxford University Press. pp 435–459.

Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK, (2005): Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 25:11730–11737.

Egner T (2011): Right ventrolateral prefrontal cortex mediates individual differences in conflict-driven cognitive control. J Cogn Neurosci 23:3903–3913.

Falkenstein M, Hohnsbein J, Hoormann J, Blanke L (1991): Effects of crossmodal divided attention on late ERP components. ii. Error processing in choice reaction tasks. Electroencephal Clin Neurophysiol 78:447–455.

Flanagan JR, Johansson RS (2003): Action plans used in action observation. Nature 424:769–771.

Friston KJ (2010): The free-energy principle: A unified brain theory? Nat Rev Neurosci 11:127–138.

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

Green DM, Swets JA (1974): Signal Detection Theory and Psychophysics. Huntington, New York: Robert E. Krieger Publishing Company.

Hikosaka O, Sesack SR, Lecourtier L, Shepard P (2008): Habenula: Crossroad between the basal ganglia and the limbic system. J Neurosci 28:11825–11829.

Holroyd CB, Krigolson OE, Baker R, Lee S, Gibson J (2009): When is an error not a prediction error? An electrophysiological investigation. Cogn Affect Behav Neurosci 9:59–70.

Keysers C, Perrett DI (2004): Demystifying social cognition: A Hebbian perspective. Trends Cogn Sci 8:501–507.

Kilner JM, Vargas C, Duval S, Blakemore S-J, Sirigu A (2004): Motor activation prior to observation of a predicted movement. Nat Neurosci 7:1299–1301.

Kilner JM, Friston KJ, Frith CD. (2007). Predictive coding: an account of the mirror neuron system. Cognitive processing, 8(3):159–166.

Koban L, Pourtois G, Vocat R, Vuilleumier P (2010): When your errors make me lose or win: Event-related potentials to observed errors of cooperators and competitors. Soc Neurosci 5:360–374.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009): Circular analysis in systems neuroscience: The dangers of double dipping. Nat Neurosci 12:535–540.

Lohmann G, Mueller K, Bosch V, Mentzel H, Hessler S, Chen L, Zysset S, von Cramon DY (2001): Lipsia—A new software system for the evaluation of functional magnetic resonance images of the human brain. Comput Med Imaging Graph 25:449–457.

Luce RD (2003): Whatever happened to information theory in psychology? Rev Gen Psychol 7:183–188.

Matsumoto M, Hikosaka O (2007): Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447: 1111–1115.

Matsumoto M, Hikosaka O (2009): Representation of negative motivational value in the primate lateral habenula. Nat Neurosci 12:77–84.

Miltner WHR, Brauer J, Hecht H, Trippe R, Coles MGH (2004): Parallel brain activity for self-generated and observed errors. In: Ullsperger M, Falkenstein M, editors. Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring. Leipzig: MPI of Cognitive Neuroscience. pp 124–129.

Nee DE, Brown JW. (2012). Dissociable FrontalStriatal and Frontal-Parietal Networks Involved in Updating Hierarchical Contexts in Working Memory. Cerebral Cortex. doi: 10.1093/cercor/bhs194.

Nichols T, Brett M, Andersson J, Wager T, Poline J-B (2005): Valid conjunction inference with the minimum statistic. NeuroImage 25:653–660.

Nichols TE, Holmes AP (2002): Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum Brain Map 15:1–25.

Nieuwenhuis S, Forstmann BU, Wagenmakers E-J (2011): Erroneous analyses of interactions in neuroscience: A problem of significance. Nat Neurosci 14:1105–1107.

Oliveira FT, McDonald JJ, Goodman D (2007): Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. J Cogn Neurosci 19:1994–2004.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004): The role of the medial frontal cortex in cognitive control. Science 306:443–447.

Rizzolatti G, Sinigaglia C (2010): The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. Nat Rev Neurosci 11:264–274.

Schie HTV, Mars RB, Coles MGH, Bekkering H (2004): Modulation of activity in medial frontal and motor cortices during error observation. Nat Neurosci 7:549–554.

Schiffer AM, Schubotz RI (2011): Caudate nucleus signals for breaches of expectation in a movement observation paradigm. Front Hum Neurosci 5:38.

Schiffer A-M, Ahlheim C, Wurm MF, Schubotz RI (2012): Surprised at all the entropy: Hippocampal, Caudate and Midbrain contributions to learning from prediction errors. PLoS One 7:e36445, doi:10.1371/journal.pone.0036445.

Schubotz RI (2007): Prediction of external events with our motor system: Towards a new framework. Trends Cogn Sci 11:211–218.

Schubotz RI, von Cramon DY (2004): Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. J Neurosci 24:5467–5474.

Schubotz RI, Korb FM, Schiffer AM, Stadler W, von Cramon DY (2012): The fraction of an action is more than a movement: Neural signatures of event segmentation in fMRI. NeuroImage 61:1195–1205.

Talairach J, Tournoux P (1988): Co-planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging. New York: Thieme.

Ullsperger M, Volz KG, von Cramon DY (2004): A common neural system signaling the need for behavioral changes. Trends Cogn Sci 8:445–447.

Volz KG, Schubotz RI, von Cramon DY (2003): Predicting events of varying probability: Uncertainty investigated by fMRI. NeuroImage 19:271–280.

Volz KG, Schubotz RI, von Cramon DY (2004): Why am I unsure? Internal and external attributions of uncertainty dissociated by fMRI. NeuroImage 21:848–857.

Volz KG, Schubotz RI, von Cramon DY (2005): Variants of uncertainty in decision-making and their neural correlates. Brain Res Bull 67:403–412.

Wessel JR, Danielmeier C, Morton JB, Ullsperger M (2012): Surprise and error: Common neuronal architecture for the processing of errors and novelty. J Neurosci 32:7528–7537.

Worsley KJ, Friston KJ (1995): Analysis of FMRI time series revisited—Again. NeuroImage 2:173–181.

Yoshida K, Saito N, Iriki A, Isoda M (2012): Social error monitoring in macaque frontal cortex. Nat Neurosci Adv. Online publication. URL http://dx.doi.org/10.1038/nn.3180.

Yu R, Zhou X (2006): Brain responses to outcomes of one's own and other's performance in a gambling task. Neuroreport 17:1747–1751.

Zacks JM, Kurby CA, Eisenberg ML, Haroutunian N (2011): Prediction error associated with the perceptual segmentation of naturalistic events. J Cogn Neurosci 23:4057–4066. Matlab algorithm: http://imaging.mrc-cbu.cam.ac.uk/downloads/MNI2tal/mni2tal.m.