

The role of trait empathy in the processing of observed actions in a false-belief task

Christian Bellebaum,¹ Marta Ghio,¹ Marie Wollmer,¹ Benjamin Weismüller,¹ and Patrizia Thoma²

¹Institute of Experimental Psychology, Heinrich-Heine University Düsseldorf, 40225 Düsseldorf, Germany, and

²Faculty of Psychology, Clinical Neuropsychology, Neuropsychological Therapy Centre, Ruhr University Bochum, 44780 Bochum, Germany

Correspondence should be addressed to Christian Bellebaum, Institute of Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, Building 23.03, Room Number 00.92, 40225 Düsseldorf, Germany. E-mail: christian.bellebaum@hhu.de

Abstract

Empathic brain responses are characterized by overlapping activations between active experience and observation of an emotion in another person, with the pattern for observation being modulated by trait empathy. Also for self-performed and observed errors, similar brain activity has been described, but findings concerning the role of empathy are mixed. We hypothesized that trait empathy modulates the processing of observed responses if expectations concerning the response are based on the beliefs of the observed person. In the present study, we utilized a false-belief task in which observed person's and observer's task-related knowledge were dissociated and errors and correct responses could be expected or unexpected. While theta power was generally modulated by the expectancy of the observed response, a negative mediofrontal event-related potential (ERP) component was more pronounced for unexpected observed actions only in participants with higher trait empathy (assessed by the Empathy Quotient), as revealed by linear mixed effects analyses. Cognitive and affective empathy, assessed by the Interpersonal Reactivity Index, were not significantly related to the ERP component. The results suggest that trait empathy can facilitate the generation of predictions and thereby modulate specific aspects of the processing of observed actions, while the contributions of specific empathy components remain unclear.

Key words: error; expectancy; empathy; theta power; oERN

Introduction

Performance monitoring plays an important role in the adaptation of behavior (Ullsperger *et al.*, 2014). A mediofrontal event-related potential (ERP) component, the error (related) negativity (Ne or ERN; Falkenstein *et al.*, 1991; Gehring *et al.*, 1993), has been described as a neural correlate of error processing. It evolves within 100 ms after an error response and is generated in the anterior cingulate cortex (ACC; Dehaene *et al.*, 1994). According to the reinforcement learning theory of the ERN (Holroyd and Coles, 2002), dopamine (DA) neurons disinhibit neurons in the ACC when an unexpected error occurs, giving rise to the ERN or, in the case of error feedback, the feedback-related negativity

(FRN; Miltner *et al.*, 1997; Gehring and Willoughby, 2002). For both error and feedback processing, corresponding ERP components have been identified when participants observe the behavior of others. Observed errors are associated with an observer ERN (oERN; Miltner *et al.*, 2004; van Schie *et al.*, 2004; Koban and Pourtois, 2014), negative feedback given to an observed person elicits an observer FRN (oFRN; Yu and Zhou, 2006; Bellebaum *et al.*, 2010), and sources in the ACC contribute to both components (Miltner *et al.*, 2004; Koban *et al.*, 2012).

While such overlapping brain activations for the experience of emotional states and the observation of these states in

Received: 10 December 2018; Revised: 11 December 2019; Accepted: 14 January 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

others are seen as a hallmark of empathic brain responses (Bernhardt and Singer, 2012; Lockwood, 2016), the relation between the electrophysiological indices of the monitoring of observed performance and trait empathy is mixed (Fukushima and Hiraki, 2006, 2009; Koban et al., 2010, 2012; Kobza et al., 2011; Thoma and Bellebaum, 2012; Rak et al., 2013; Thoma et al., 2015; Mothes et al., 2016). To gain new insights into the determinants of this relationship is the main motivation for the present study.

It has been argued that the processing of observed responses and their outcomes depends on a number of factors such as task and context (Koban and Pourtois, 2014), involving multiple cognitive and affective processes, only some of which may be affected by empathy. For example, concerning the outcomes of others' actions, Marco-Pallarés et al. (2010) suggested that the neural response is determined by two processes: an 'empathic' one related to the consequences for the observed person and a self-centered one related to the consequences for the observer. The empathic process, as reflected by the oERN, is stronger in a cooperative context and for higher levels of perspective taking (Koban et al., 2010, 2012), which is considered a cognitive component of empathy (Shamay-Tsoory, 2011). Moreover, brain activity in the ACC has been shown to code for other's rewards (Apps et al., 2013; Chang et al., 2013), and in individuals with high levels of emotion contagion, which reflects a basic empathic process, the gyral ACC specifically predicted others' rewards (Lockwood et al., 2015). While empathic processes thus appear to facilitate reward prediction from another's perspective, a strong tendency to focus on others can also lead to impairments in learning from other's actions for one's own benefit (Kobza et al., 2011).

Predictions also play an important role when simple actions are observed. As revealed by predictive eye movements, observers generate hypotheses about upcoming actions (Flanagan and Johansson, 2003; Ambrosini et al., 2011; Costantini et al., 2014; Donnarumma et al., 2017). In line with this assumption, we found that a mediofrontal negative ERP component in the time window of the oERN was most pronounced for unexpected observed actions (Kobza and Bellebaum, 2013), possibly reflecting the processing of an action prediction error (Burke et al., 2010). Similar to Apps et al. (2013) who investigated the processing of other's outcome prediction errors, we used a false-belief task to dissociate task-related knowledge in the observer and the observed person (Kobza and Bellebaum, 2013). False-belief tasks are thought to require theory of mind and perspective taking (e.g. Ferguson et al., 2015; Birch et al., 2017; Rubio-Fernández, 2017). As unexpected correct responses were errors from the observed person's perspective in our task (Kobza and Bellebaum, 2013), it is conceivable that perspective taking enabled our participants to process correct responses as vicarious errors in the false-belief condition. On the other hand, findings by Lockwood et al. (2015) hint at the importance of affective empathic processes for ACC-driven prediction processes.

The present study thus aimed to elucidate the relationship between trait empathy and action observation. We hypothesized that both affective and cognitive empathy contribute to the generation of predictions about other's upcoming actions and thus to the processing of these actions in a false-belief task. In addition to the mediofrontal ERP component, we also analyzed response-locked theta power in the frequency domain. Theta has been shown to be enhanced for the processing of own and observed errors (e.g. Cohen, 2011; Pezzetta et al., 2018) and is

believed to reflect general cognitive control processes during performance monitoring (Cavanagh and Frank, 2014).

Materials and methods

Participants

Fifty healthy young women with a mean age of 23.8 years ($SD = 3.0$, range 19–32) took part in this study. Power calculations for mixed model analyses (see below) are not straightforward (Mathieu et al., 2012), and there was no available evidence from studies with comparable design, research question and analysis procedure on which we could base an educated guess of the effect size. Therefore, the sample size was only approximated based on the previous studies from the performance monitoring or decision-making domain, in which the effect of a categorical (instead of continuous, see below) between-subjects factor on two within-subjects factors was tested. A sample size of 20–30 participants per group, and thus 40–60 in total, is usual for such studies (e.g. Bellebaum and Colosio, 2014; Luo et al., 2014; Barker et al., 2015), and we thus decided to test 50 participants. Out of these participants, 46 reported to be right-handed and four left-handed. Exclusion criteria were a history of psychiatric or neurological disease and any regular medication affecting the central nervous system, as assessed by self-report. All participants had normal or corrected to normal vision and received a reimbursement of 15€ for their participation in the study. The procedures were in accordance with the Declaration of Helsinki, and the study was approved by the ethics committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University, Düsseldorf, Germany.

Experimental task

We applied a variant of an experimental task that we used in a previous, related study (see Kobza and Bellebaum, 2013). The task involves the observation of another person's choice actions on a computer screen. The study participants were asked to observe the actions carefully. The observed person in the present study was always a woman, and she was introduced with a picture and a name on the screen. In fact, however, the observed person was only virtual, and the performance that the participant observed was a pre-determined one. Stimulus timing and response recording were controlled with the software Presentation (Neurobehavioral Systems Inc). According to the instructions, the observed person engaged in the Two Shell Game, where she was asked to guess under which of two 'shells' a 'ball' was hidden. Only the hand of the observed person was visible, which held a joystick in the lower center of the screen (see Figure 1). The observer could always see the ball in one of the shells, but she was told that the observed person could not see the ball from her perspective. After the shells rapidly changed positions via rotating movements, the observed person indicated with a joystick movement on which side she thought the ball was hidden (left or right). As the rotating movements were easy to follow, the observer should expect that the observed person would choose the correct position. Importantly, the observers were instructed that the observed person was 'tricked' on some trials, where the ball swapped shells during the rotation movements ('trick' condition, as opposed to the 'no trick' condition). The swapping could be detected by the observers, but according to the instruction, not by the observed person, so that the observers should expect an error by the observed person. The task can thus

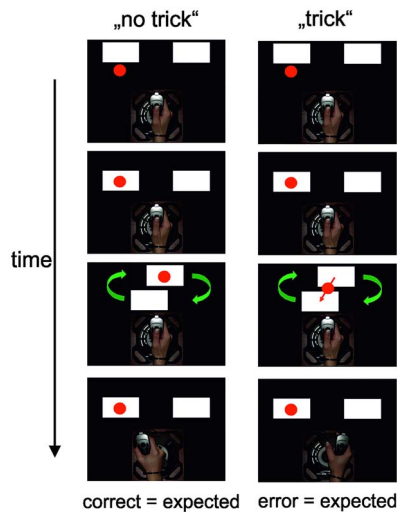


Fig. 1. Sequence of events in the ‘no trick’ and ‘trick’ trials. The observers saw the ball and its location in one of the ‘shells’ during the whole trial. In the ‘trick’ condition, the ball was swapped between ‘shells’, which could be seen by the observers, but not by the observed person, according to the instruction.

be considered as a false-belief task, because the observer had privileged task-related knowledge.

In total, the task consisted of 468 trials. In 48 randomly interspersed trials (24 for the ‘trick’ and ‘no trick’ conditions, respectively), the 2 shells stayed on the screen for 400 ms after the rotation movements were completed and no response was observed. The observer participants were then prompted to indicate where they thought the observed person would move the joystick by pressing the left or right CTRL key of a computer keyboard. With these trials, the expectancy of the observers concerning a particular response could be assessed.

From the remaining 420 (non-prompt) trials, 50% were ‘no trick’ and 50% were ‘trick’ trials. For each trial type, the choices by the observed person were as often objectively correct as they were erroneous. Thus, there were 105 trials for each of the 4 conditions ‘trick’ correct, ‘trick’ error, ‘no trick’ correct and ‘no trick’ error.

EEG recording

While the participants completed the experimental task, EEG data were recorded at a sampling rate of 1000 Hz from 29 scalp sites using active silver/silver-chloride electrodes attached in accordance with the international 10-10 system. Electrode positions were F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2 and PO10, with an electrode at FCz serving as online reference. For recording, a 32-channel actiCAP electrode cap (ActiCAP; Brain Products GmbH, Germany) was used. Horizontal eye movements were assessed with an electrode placed next to the outer canthus of the left eye. Blinks and vertical eye movements were recorded by an electrode positioned at Fp1 and thus above the left eye. The BrainVision Recorder software, version 1.20 (Brain Products, Munich, Germany), was used for the recording of the data. The impedances were lower than 10 k Ω .

Assessment of empathy

Trait empathy was assessed using the German version of the Cambridge Empathy Scale, a self-report questionnaire that

yields the so-called empathy quotient (EQ; Baron-Cohen and Wheelwright, 2004), based on the summed scores of the 40 relevant items. In order to explore if specific empathy components are particularly important for the processing of observed actions, we additionally applied a German short version (Paulus, 2007) of the Interpersonal Reactivity Index (IRI; Davis, 1980, 1983), which consists of four subscales with four items: fantasy, perspective taking, empathic concern and personal distress. For each participant, we added up the fantasy and perspective taking subscores as a measure of cognitive empathy and the empathic concern and personal distress subscores as a measure of affective empathy (see e.g. Shamay-Tsoory et al., 2009).

Data analysis

Behavioral data. The responses of the observer participants in the prompt trials were analyzed by determining a) the percentage of trials in which correct responses were expected in the ‘no trick’ condition and b) the percentage of trials in which error responses were expected in the ‘trick’ condition. Concerning the statistical analysis, we first aimed to verify that our experimental conditions successfully induced specific expectations of a certain observed response. For this purpose, one sample t-tests were used to compare the expectancy measures derived from each condition (see above) with a value of 50%, which would indicate no specific expectation concerning the type of response (error or correct). Then the expectancy measures were compared between conditions by applying a paired t-test (applying the software IBM SPSS statistics, version 23). No further analyses of a potential relationship between expectancy and empathy were performed, because in both the ‘trick’ and the ‘no trick’ conditions, the expectancy values were very high, with low interindividual variability and a strongly left-skewed, non-normal distribution.

EEG data. The analysis of the EEG data focused on an ERP component time-locked to the observed response that was described in a previous study by our group (Kobza and Bellebaum, 2013). In addition, the induced theta power following the observed responses was analyzed (see the supplementary materials). BrainVision Analyzer software (version 2.1, Brain Products, Munich, Germany) and MATLAB (MathWorks, Natick, Massachusetts) were used for the analyses.

At first, the data were re-referenced to linked mastoids, and the data for electrode FCz were reconstructed. Then a global direct detrend procedure was applied to correct for drifts: Then a 20 Hz low-pass and a 0.5 Hz high-pass filter were applied to the EEG data. An independent component analysis (ICA) was applied to a 240 s excerpt of each participant’s EEG data (starting 120 s after the start of the experiment) from the 29 scalp recording sites in order to identify (a) component(s) reflecting eye blinks. The 240 s interval was long enough for several eye blinks to occur in each participant. ICA components were considered to represent blink artifacts if they represented a symmetrical, frontally pronounced positivity in the signal. A back transformation was then performed to remove the identified component(s) from all the raw data, thus correcting for eye blinks. Epochs of 800 ms length were created, starting from 200 ms before to 600 ms after the observed joystick movement. A baseline correction relative to the 200 ms before the observed response was applied. Segments with voltage steps of more than 50 μ V/ms, amplitude differences between the maximum and minimum value of more than 100 μ V or less than 0.1 μ V and with absolute amplitude values above 100 μ V were excluded. In

the final step, the average ERPs were calculated separately for the four conditions 'trick' correct, 'trick' error, 'no trick' correct and 'no trick' error at each electrode. In accordance with our previous study (Kobza and Bellebaum, 2013), we quantified a negativity in the ERP calculating a peak-to-peak amplitude. As the negativity was most pronounced in a frontocentral electrode cluster, the signal was pooled across the electrodes Fz, FC1, FCz, FC2 and Cz. Then the maximum negative peak in the time window between 250 and 420 ms after the observed response was determined in the pooled signal. The maximum positive amplitude in the preceding time window between 130 ms, and the negative peak was subtracted from the negative peak.

To examine the relationship between trait empathy and neural indices of action observation, we applied linear mixed effects (LME) analyses. This approach allowed us to define models in which one or more continuous predictor variable(s) such as the different empathy measures (see above) and the two categorical factors trial type and choice accuracy interact to predict an outcome variable. Specifically, we conducted two separate LME analyses in order to explore in how far different aspects of empathy affect the processing of the observed actions, alone or in interaction with the categorical factors. In the first LME analysis, we specified a model that included as fixed-effect predictors the continuous factor EQ sum score (mean-centered) and the categorical factors trial type (recoded as +1 = trick, -1 = no trick) and choice accuracy (recoded as +1 = correct, -1 = error), as well as their interactions. The model included the amplitude of the negative mediofrontal ERP component as a dependent variable and the participants as a random-effect factor.

Then we explored if affective and/or cognitive empathy affect(s) the processing of other's actions. In a second LME analysis, we thus defined a model including the same categorical predictor variables and random-effect factors as described above, but we specified the cognitive and the affective empathy scores derived from the IRI as two continuous factors (mean-centered). These two measures are considered to reflect independent aspects of empathy (e.g. Shamay-Tsoory et al., 2009), and they did not correlate significantly in our sample of participants ($r = 0.106$; $P = 0.465$). The latter is a pre-requisite for entering two continuous fixed-effect predictors into one LME analysis. Both analyses were conducted by using the `lme4` statistical package (version 1.1-18) in the R environment (version 3.5.1). Following the approach suggested by Luke (2017), in each LME analysis, the model was estimated by using a restricted maximum likelihood approach, and the R package `lmerTest` (version 3.0-1, Kuznetsova et al., 2017) was applied for evaluating significance by using Satterthwaite approximation for the degrees of freedom. This approach has been shown to produce acceptable Type I error rates and to be less sensitive to the sample size compared to other approaches, without a noticeable loss of power (Luke, 2017). As we conducted two LME analyses involving different, but related measures of empathy (the EQ sum score correlated significantly with cognitive empathy, $r = 0.449$; $P = 0.001$, but not with affective empathy, $r = 0.061$; $P = 0.673$), we applied a Bonferroni correction to the threshold of statistical significance and set it to $P < 0.025$.

One participant had to be excluded from the analysis of the ERP data because no peaks could be detected in or near the time windows for the analysis (see above). Furthermore, we tested the data entered into each of the two LME analyses for statistical outliers on the participant level by using the R package `influenceME` (Nieuwenhuis et al., 2012). Five participants with a Cook's distance value above the cutoff calculated as $4/(n-k-1)$, with n = number of participants and k = number of factors, were

identified as influential data points in at least one of the tested models and therefore excluded from all analyses for the frontocentral negativity in the ERP. Thus, both LME analyses for the negative mediofrontal ERP component were conducted with a sample of 44 participants. Corresponding LME analyses with the same models, factors and threshold for statistical significance were conducted for theta power (see supplementary materials).

Results

Behavioral data

On average, the participants expected a correct response by the observed person in 84.5% (SD = 16.8%) of the prompt trials of the 'no trick' condition and an error response in 88.4% (SD = 16.2%) of the prompt trials of the 'trick' condition. Both values differed significantly from chance (50%), indicating that the participants indeed expected particular responses in both conditions, $t(49) = 14.513$, $P < 0.001$, $d = 2.054$ ('no trick') and $t(49) = 16.754$, $P < 0.001$, $d = 2.370$ ('trick'). The strength of the expectations for error responses in the 'trick' condition and for correct responses in the 'no trick' condition did not differ significantly ($P = 0.106$).

EEG data

Figure 2 shows the grand average waveforms (based on the 44 participants entering the LME analyses) of the ERPs following observed correct and error responses in the 'no trick' and 'trick' conditions, as well as the topographies of the mediofrontal negative ERP component for error responses in the 'no trick' condition and correct responses in the 'trick' condition. The topographies show that the negativity is largest at the five mediofrontal electrode sites for which the data were pooled.

The first LME analysis tested the effects of the continuous predictor variable EQ sum score and the two categorical factors trial type and choice accuracy. This analysis revealed that all main effects as well as the choice accuracy by EQ sum score and the trial type by EQ sum score interactions were not significant (all $P > 0.077$; for parameter-specific t -tests see also Table 1). The interaction between trial type and choice accuracy approached significance, $F(1,126) = 4.964$, $P = 0.028$ (see Methods section for information on Bonferroni correction of the threshold for statistical significance). On average, the amplitude of the negative ERP component for errors in the 'no trick' condition was $-3.91 \mu\text{V}$ (SD = $1.63 \mu\text{V}$), while for correct responses, the amplitude was $-3.33 \mu\text{V}$ (SD = $1.57 \mu\text{V}$). In the 'trick' condition, the pattern was reversed, with slightly higher (i.e. more negative) amplitudes for correct (mean = $-3.64 \mu\text{V}$, SD = $1.77 \mu\text{V}$) than error responses (mean = $-3.58 \mu\text{V}$, SD = 1.81). As the two-way interaction was not significant, we did not further resolve it. However, we found a significant three-way interaction, $F(1,126) = 11.609$, $P < 0.001$ (see also Figure 3). In order to resolve the three-way interaction, we applied simple slope analyses (Liu et al., 2017) by using the R package `jtool` (version 1.1.1). More specifically, we computed a simple slope analysis, in which we determined the slope of the negative ERP component on choice accuracy at fixed cutoff values of EQ sum score (centered mean ± 1 SD; see Liu et al., 2017) for each level of the predictor trial type (i.e. 'trick', 'no trick'). As also suggested by Figure 3, the amplitude of the negative ERP component was not affected by choice accuracy for lower EQ sum scores, neither in the 'trick' nor in the 'no trick' condition (both $P > 0.086$). For higher EQ sum scores, the amplitude was significantly modulated by choice accuracy in both trial type conditions, but in opposite directions.

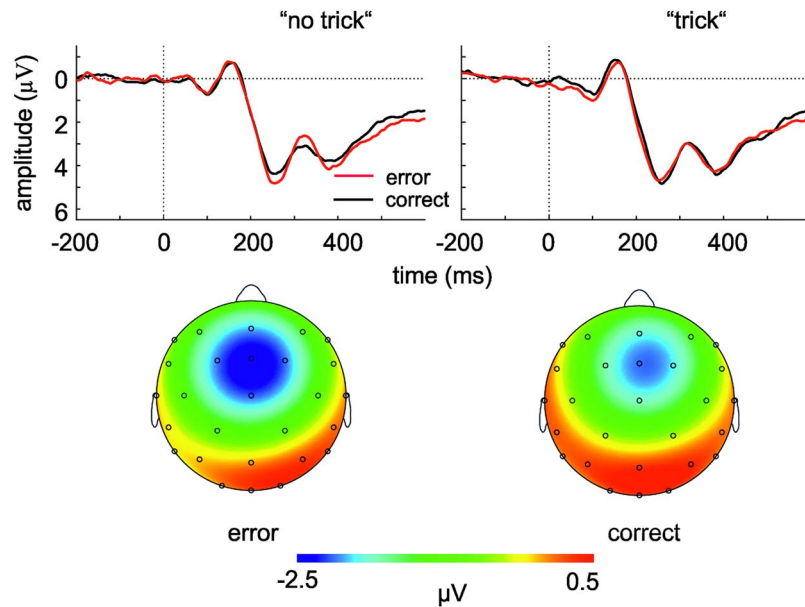


Fig. 2. In the upper panel, the grand average ERPs time-locked to erroneous and correct responses in the 'no trick' and 'trick' conditions are displayed (pooled for the electrode positions Fz, FC1, FCz, FC2 and Cz; see Methods section). The lower panel shows the topographies of the ERPs for error responses in the 'no trick' condition and correct responses in the 'trick' condition, relative to the preceding positive peak.

Table 1. Summary of the estimated linear mixed effects model, with parameter-specific t-tests for the main effects of trial type, choice accuracy and EQ sum score and their interactions

	Estimate	Std. error	df	t-value	P-value	CI 2.5%	CI 97.5%
(Intercept)	-3.613	0.222	42	-16.246	<0.001	-4.049	-3.177
Trial type	0.005	0.073	126	0.066	0.948	-0.138	0.148
Choice accuracy	0.130	0.073	126	1.783	0.077	-0.013	0.273
EQ sum score	0.005	0.024	42	0.216	0.830	-0.042	0.053
Trial type × choice accuracy	-0.162	0.073	126	-2.228	0.028	-0.305	-0.020
Trial type × EQ sum score	-0.010	0.008	126	-1.306	0.194	-0.026	0.005
Choice × EQ sum score	-0.004	0.008	126	-0.487	0.627	-0.019	0.012
Trial type × choice accuracy × EQ sum score	-0.027	0.008	126	-3.407	<0.001	-0.042	-0.011

In the 'trick' condition, we observed a larger negative ERP amplitude for correct than error responses ($b = -0.317$, $P = 0.032$), whereas in the 'no trick' condition, a larger negative ERP amplitude for error than correct responses emerged ($b = 0.506$, $P < 0.001$).

The second LME analysis tested the interaction between trial type, choice accuracy, cognitive and affective empathy. This analysis revealed a consistent pattern of results, with the interaction between trial type and choice accuracy again approaching significance ($P = 0.035$, see above). Concerning the effects of empathy, however, all main and interaction effects including the factors cognitive or affective empathy were not significant (all $P > 0.086$; see Table 2 for a summary of the estimated mixed-effects model, with parameter-specific t-tests for all the effects in this analysis). For the results concerning theta power, see the supplementary materials.

Discussion

The relationship between empathy and observational performance monitoring is the issue of an ongoing debate. One reason for the inconsistency in the findings (Thoma and Bellebaum, 2012; Amiruddin et al., 2017) might be that empathy is relevant for only some cognitive and affective processes during

action observation and that the processes required differed between previous studies. In the present study, we assumed that trait empathy would modulate the processing of observed responses in a task where participants were asked to predict the responses of an observed person under true- and false-belief conditions (Kobza and Bellebaum, 2013). Using LME analyses, we found a relationship between a general measure of empathy and the processing of observed actions, as reflected in a negative mediofrontal ERP component, but not theta power (see supplementary materials). Participants with higher but not lower empathy showed larger ERP amplitudes to correct than error responses in 'trick' trials, where correct responses were unexpected due to a false belief of the observed person. On the contrary, higher but not lower empathy was associated with stronger responses to (unexpected) observed errors than (expected) correct responses in the true-belief ('no trick') condition. Affective or cognitive empathy did not significantly modulate the processing of observed responses.

Expectancy, empathy and the processing of observed actions

In our previous study involving the same experimental paradigm, we have interpreted the negative ERP component

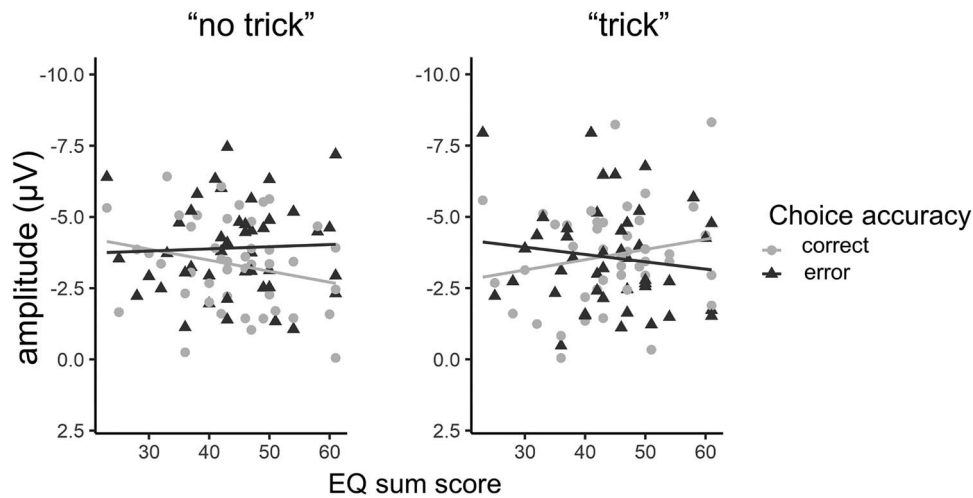


Fig. 3. Single participants' amplitudes (μV) of the negative mediofrontal ERP component depending on the EQ sum score and the choice accuracy of the observed response (correct in grey, error in black) in the 'no trick' (left panel) and 'trick' (right panel) condition.

Table 2. Summary of the linear mixed effect model for the negative mediofrontal ERP component including the effects of trial type, choice accuracy, cognitive (Cog) and affective (Aff) empathy and their interactions

	Estimate	Std. error	df	t-value	P-value	CI 2.5%	CI 97.5%
(Intercept)	-3.607	0.224	40	-16.137	<2e - 16	-4.045	-3.169
Trial	0.008	0.076	120	0.108	0.915	-0.142	0.158
Choice	0.132	0.076	120	1.729	0.086	-0.018	0.282
Aff	-0.061	0.055	40	-1.105	0.276	-0.168	0.047
Cog	0.016	0.061	40	0.267	0.791	-0.104	0.137
Trial \times choice	-0.163	0.076	120	-2.129	0.035	-0.313	-0.013
Trial \times Aff	-0.003	0.019	120	-0.146	0.884	-0.039	0.034
Choice \times Aff	0.011	0.019	120	0.597	0.551	-0.026	0.048
Trial \times Cog	-0.026	0.021	120	-1.228	0.222	-0.067	0.015
Choice \times Cog	-0.006	0.021	120	-0.291	0.771	-0.047	0.035
Aff \times Cog	-0.013	0.017	40	-0.748	0.459	-0.046	0.021
Trial \times choice \times Aff	-0.022	0.019	120	-1.186	0.238	-0.059	0.014
Trial \times choice \times Cog	-0.022	0.021	120	-1.068	0.288	-0.064	0.019
Trial \times Aff \times Cog	-0.007	0.006	120	-1.253	0.213	-0.019	0.004
Choice \times Aff \times Cog	-0.005	0.006	120	-0.815	0.417	-0.016	0.007
Trial \times choice \times Aff	0.001	0.006	120	0.127	0.899	-0.011	0.012
Aff \times Cog							

as an indicator of the processing of unexpected observed actions (Kobza and Bellebaum, 2013). This interpretation is in line with theoretical accounts of ACC function as well as ERP and neuroimaging evidence linking it to the processing of unexpected events in general, irrespective of their valence (Jessup et al., 2010; Alexander and Brown, 2011; Ferdinand et al., 2012; Schiffer et al., 2014). It also complies with the firing pattern of single neurons in the monkey ACC, which have been reported to specifically predict another monkey's unknown decision (Haroush and Williams, 2015).

The present study provides new insights into the role of empathy in the monitoring of observed responses. We found a relationship between the mediofrontal ERP component and the EQ sum score, which is a general measure of empathy and comprises the factors 'cognitive empathy', 'emotional reactivity' and 'social skills' (Lawrence et al., 2004). Only in highly empathic individuals, the observation of more unexpected responses yielded larger negative amplitudes than that of expected responses, in both the false- and true-belief conditions. This finding might

suggest that empathy facilitates the generation of expectations concerning upcoming responses by an observed person, leading to stronger action prediction errors and, consequently, larger mediofrontal ERPs. A similar interpretation was provided by Ferguson et al. (2015) who used EEG to examine the processing of target words that described where another person searched for an object. The N400 ERP component in highly empathic individuals (also quantified as the EQ sum score) reflected the degree to which the target word was processed from the described person's perspective in a false-belief condition. The authors suggested that empathy helped to overcome an 'egocentric bias' in the evaluation of the other person's action. A similar process might have occurred in the present study.

At the same time, this interpretation has to be treated with caution as our empathy measures were not consistent in pointing in this direction. We did not find evidence for a specific role of cognitive or affective empathy, as assessed by the IRI, concerning the processing of observed actions. This might seem surprising, given that cognitive and affective aspects of empathy have been

linked to performance in false-belief tasks (Birch et al., 2017; Rubio-Fernández, 2017) and to ACC-driven prediction processes, respectively (Lockwood et al., 2015).

On the one hand, the negative finding for the role of cognitive and affective empathy might mean that these processes alone do not modulate the processing of observed responses. On the other hand, there is some controversy in the literature about whether the IRI subscales ‘fantasy’ (contributing to the cognitive empathy score) and ‘personal distress’ (contributing to the affective empathy score) should be regarded as aspects of empathy (see Sindermann et al., 2019 for a recent summary of this discussion). As the authors have also highlighted, various studies across different countries have reported that only the ‘perspective taking’ and ‘empathic concern’ IRI subscales show strong correlations with the EQ. This might also be a reason why, in our analyses, we did not find any effects for the measures derived from the IRI, where we aggregated the ‘fantasy’ and ‘perspective taking’ scores into a ‘cognitive empathy’ score and the ‘empathic concern’ and ‘personal distress’ scores into an ‘affective empathy’ score. Furthermore, as Koller and Lamm (2015) point out, the IRI subscales apart from ‘empathic concern’ might not even assess homogeneous constructs for themselves. For instance, both the ‘fantasy’ and the ‘perspective taking’ subscales might assess two subdimensions each: The ‘perspective taking’ subscale might encompass one more cognitive comparison component (understanding both sides of a story) and another more simulation-based component (putting oneself into another person’s shoes). Only the latter might be relevant for the interpretation of the present findings. Also, Koller and Lamm (2015) stress that the IRI subscales might comprise too few items (four each) to assess the underlying constructs in a stable manner. In contrast to this, the EQ comprises 40 items that assess the underlying construct of empathy in a fairly unidimensional manner. All of these factors might have contributed to the discrepant findings we have found for the different empathy measures.

Study limitations

The main limitation of the present study is that the observation situation was not real. While we did not assess whether our participants believed to observe a real person, our experimental manipulation with the ‘trick’ and ‘no trick’ conditions was successful in inducing different expectations concerning the accuracy of the observed response. Furthermore, an exploratory analysis revealed that the expectations did not change significantly during the experiment, suggesting that the artificial observation situation and the ‘unnatural’ equal frequencies of errors and correct responses did not strongly affect the participants’ beliefs.

A related point refers to the lack of a non-social control condition. While also other studies using false-belief tasks are lacking such a condition (e.g. Ferguson et al., 2015), we acknowledge that our interpretation of the role of empathy would be strengthened if we showed that empathy modulates the processing of observed actions only in social situations.

Conclusion

In summary, we showed that in a task with the specific requirement to predict an observed person’s response under true- and false-belief conditions, trait empathy, at least as measured by the unidimensional EQ, affects the processing

of observed responses, as reflected by a mediofrontal ERP component. Although our main interpretation is that general empathy can facilitate the generation of predictions concerning upcoming responses by others based on the other person’s assumed mental state, the fact that empathy is not a unitary concept which is operationalized very differently by distinct questionnaire measures also has to be kept in mind on a cautionary note.

Supplementary data

Supplementary Material is available at SOCAFN online

Conflict of interest

The authors declare no conflict of interest.

References

- Alexander, W.H., Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*, 1338–44.
- Ambrosini, E., Costantini, M., Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*, 1437–42.
- Amiruddin, A., Fueggle, S.N., Nguyen, A.T., Gignac, G.E., Clunies-Ross, K.L., Fox, A.M. (2017). Error monitoring and empathy: explorations within a neurophysiological context. *Psychophysiology*, *54*, 864–73.
- Apps, M.A.J., Green, R., Ramnani, N. (2013). Reinforcement learning signals in the anterior cingulate cortex code for others’ false beliefs. *Neuro Image*, *64*, 1–9.
- Barker, T.V., Troller-Renfree, S., Pine, D.S., Fox, N.A. (2015). Individual differences in social anxiety affect the salience of errors in social contexts. *Cognitive, Affective, & Behavioral Neuroscience*, *15*, 723–35.
- Baron-Cohen, S., Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163–75.
- Bellebaum, C., Colosio, M. (2014). From feedback- to response-based performance monitoring in active and observational learning. *Journal of Cognitive Neuroscience*, *26*, 2111–27.
- Bellebaum, C., Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *The European Journal of Neuroscience*, *27*, 1823–35.
- Bellebaum, C., Kobza, S., Thiele, S., Daum, I. (2010). It was not MY fault: event-related brain potentials in active and observational learning from feedback. *Cerebral Cortex*, *20*, 2874–83.
- Bernhardt, B.C., Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience*, *35*, 1–23.
- Birch, S.A., Li, V., Haddock, T., et al. (2017). Perspectives on perspective taking: how children think about the minds of others. *Advances in Child Development and Behavior*, *52*, 185–226.
- Burke, C.J., Tobler, P.N., Baddeley, M., Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 14431–6.
- Cavanagh, J.F., Frank, M.J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*, 414–21.
- Cavanagh, J.F., Frank, M.J., Klein, T.J., Allen, J.J.B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuro Image*, *49*, 3198–209.
- Chang, S.W.C., Gariépy, J.-F., Platt, M.L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nature Neuroscience*, *16*, 243–50.

- Cohen, M.X. (2011). Error-related medial frontal theta activity predicts cingulate-related structural connectivity. *Neuro Image*, *55*, 1373–83.
- Costantini, M., Ambrosini, E., Cardellicchio, P., Sinigaglia, C. (2014). How your hand drives my eyes. *Social Cognitive and Affective Neuroscience*, *9*, 705–11.
- Davis, M.H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.
- Davis, M.H. (1983). Measuring individual-differences in empathy-evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–26.
- Dehaene, S., Posner, M.I., Tucker, D.M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, *5*, 303–5.
- Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., Pezulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45–60.
- 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. *Electroencephalography and Clinical Neurophysiology*, *78*, 447–55.
- Ferdinand, N.K., Mecklinger, A., Kray, J., Gehring, W.J. (2012). The processing of unexpected positive response outcomes in the medial frontal cortex. *The Journal of Neuroscience*, *32*, 12087–92.
- Ferguson, H.J., Cane, J.E., Douchkov, M., Wright, D. (2015). Empathy predicts false belief reasoning ability: evidence from the N400. *Social Cognitive and Affective Neuroscience*, *10*, 848–55.
- Flanagan, J.R., Johansson, R.S. (2003). Action plans used in action observation. *Nature*, *424*, 769–71.
- Frith, C.D., Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*, R724–32.
- Fukushima, H., Hiraki, K. (2006). Perceiving an opponent's loss: gender-related differences in the medial-frontal negativity. *Social Cognitive and Affective Neuroscience*, *1*, 149–57.
- Fukushima, H., Hiraki, K. (2009). Whose loss is it? Human electrophysiological correlates of non-self reward processing. *Social Neuroscience*, *4*, 261–75.
- Gehring, W.J., Willoughby, A.R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*, 2279–82.
- Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–90.
- Haroush, K., Williams, Z.M. (2015). Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell*, *160*, 1233–45.
- Holroyd, C.B., Coles, M.G.H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.
- Janssen, D.J.C., Poljac, E., Bekkering, H. (2016). Binary sensitivity of theta activity for gain and loss when monitoring parametric prediction errors. *Social Cognitive and Affective Neuroscience*, *11*, 1280–9.
- Jessup, R.K., Busemeyer, J.R., Brown, J.W. (2010). Error effects in anterior cingulate cortex reverse when error likelihood is high. *The Journal of Neuroscience*, *30*, 3467–72.
- Koban, L., Pourtois, G. (2014). Brain systems underlying the affective and social monitoring of actions: an integrative review. *Neuroscience and Biobehavioral Reviews*, *46*(Pt 1), 71–84.
- 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. *Social Neuroscience*, *5*, 360–74.
- Koban, L., Pourtois, G., Bediou, B., Vuilleumier, P. (2012). Effects of social context and predictive relevance on action outcome monitoring. *Cognitive, Affective, & Behavioral Neuroscience*, *12*, 460–78.
- Kobza, S., Bellebaum, C. (2013). Medial frontal event-related potentials following observed actions reflect an action prediction error. *The European Journal of Neuroscience*, *37*, 1435–40.
- Kobza, S., Thoma, P., Daum, I., Bellebaum, C. (2011). The feedback-related negativity is modulated by feedback probability in observational learning. *Behavioural Brain Research*, *225*, 396–404.
- Koller, I., Lamm, C. (2015). Item response model investigation of the (German) interpersonal reactivity index empathy questionnaire implications for analyses of group differences. *European Journal of Psychological Assessment*, *31*, 211–21.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*.
- Lawrence, E.J., Shaw, P., Baker, D., Baron-Cohen, S., David, A.S. (2004). Measuring empathy: reliability and validity of the empathy quotient. *Psychological Medicine*, *34*, 911–9.
- Liu, Y., West, S.G., Levy, R., Aiken, L.S. (2017). Tests of simple slopes in multiple regression models with an interaction: comparison of four approaches. *Multivariate Behavioral Research*, *52*, 445–64.
- Lockwood, P.L. (2016). The anatomy of empathy: vicarious experience and disorders of social cognition. *Behavioural Brain Research*, *311*, 255–66.
- Lockwood, P.L., Apps, M.A.J., Roiser, J.P., Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *The Journal of Neuroscience*, *35*, 13720–7.
- Luke, S.G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–502.
- Luo, Y., Wu, T., Broster, L.S., et al. (2014). The temporal course of the influence of anxiety on fairness considerations. *Psychophysiology*, *51*, 834–42.
- Marco-Pallarés, J., Krämer, U.M., Strehl, S., Schröder, A., Münte, T.F. (2010). When decisions of others matter to me: an electrophysiological analysis. *BMC Neuroscience*, *11*, 86.
- Mathieu, J.E., Aguinis, H., Culpepper, S.A., Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*, 951–66.
- Miltner, W.H., Braun, C.H., Coles, M.G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. *Journal of Cognitive Neuroscience*, *9*, 788–98.
- Miltner, W.H.R., Brauer, J., Hecht, H., Trippe, R., Coles, M.G.H. (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.
- Mothes, H., Enge, S., Strobel, A. (2016). The interplay between feedback-related negativity and individual differences in altruistic punishment: an EEG study. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 276–88.
- Nieuwenhuis, R., te Grotenhuis, H.F., Pelzer, B.J. (2012). Influence ME: tools for detecting influential data in mixed effects models. *The R-Journal*, *4*(2), 38–47.
- Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. [The SPF (IRI) for the measure of empathy: Psychometric evaluation of the German interpersonal reactivity index]. Retrieved from <http://psydok.sulb.uni-saarland.de/volltexte/2009/2363>.

- Pezzetta, R., Nicolardi, V., Tidoni, E., Aglioti, S.M. (2018). Error, rather than its probability, elicits specific electrocortical signatures: a combined EEG-immersive virtual reality study of action observation. *Journal of Neurophysiology*, **120**, 1107–18.
- Rak, N., Bellebaum, C., Thoma, P. (2013). Empathy and feedback processing in active and observational learning. *Cognitive, Affective, & Behavioral Neuroscience*, **13**, 869–84.
- Rubio-Fernández, P. (2017). Why are bilinguals better than monolinguals at false-belief tasks? *Psychonomic Bulletin & Review*, **24**, 987–98.
- Schiffer, A.-M., Krause, K.H., Schubotz, R.I. (2014). Surprisingly correct: unexpectedness of observed actions activates the medial prefrontal cortex. *Human Brain Mapping*, **35**, 1615–29.
- Shamay-Tsoory, S.G., Aharon-Peretz, J., Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, **132**, 617–27.
- Shamay-Tsoory, S.G. (2011). The neural bases for empathy. *The Neuroscientist*, **17**, 18–24.
- Sindermann, C., Cooper, A., Montag, C. (2019). Empathy, autistic tendencies, and systemizing tendencies-relationships between standard self-report measures. *Frontiers in Psychology*, **10**, 307.
- Thoma, P., Bellebaum, C. (2012). Your error's got me feeling-how empathy relates to the electrophysiological correlates of performance monitoring. *Frontiers in Human Neuroscience*, **6**, 135.
- Thoma, P., Norra, C., Juckel, G., Suchan, B., Bellebaum, C. (2015). Performance monitoring and empathy during active and observational learning in patients with major depression. *Biological Psychology*, **109**, 222–31.
- Ullsperger, M., Danielmeier, C., Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological Reviews*, **94**, 35–79.
- van Schie, H.T., Mars, R.B., Coles, M.G.H., Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, **7**, 549–54.
- Yu, R., Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport*, **17**, 1747–51.