Biological Psychology 194 (2025) 108965

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

To err is human: Differences in performance monitoring ERPs during interactions with human co-actors and machines

Bence Neszmélyi [a,*], Roland Pfister [a,b,c,**]

[a] Department of Psychology III, University of Würzburg, Röntgenring 11, Würzburg 97070, Germany
[b] Department of Psychology, Trier University, Johanniterufer 15, Trier 54290, Germany
[c] Institute for Cognitive and Affective Neuroscience (ICAN), Trier University, Johanniterufer 15, Trier 54290, Germany

ARTICLE INFO

Keywords:
Error observation
Social performance monitoring
Human-machine interaction
Observed-error-related negativity
Observed-error positivity

ABSTRACT

In interactive tasks, agents often aim at eliciting a certain response from their partner. Not accomplishing this goal calls for adjusting behavior on the fly. Previous research suggests that such adjustments differ when interacting with a machine or with a fellow human agent. In this study, we investigated whether such differences are also reflected in event-related potentials induced by observing human and machine errors in an interactive setting. In a four-choice reaction time task, participants performed actions that were followed by regular and irregular visual effects. In different conditions, participants were led to believe that they were interacting with another human agent or with a machine so that the irregular effects were attributed either to human errors or to machine malfunctions. We compared observed-error-related negativity (oERN) and observed-error positivity (oP_E) for these two error types. The oP_E was not affected by the experimental manipulation, whereas the oERN amplitude was more pronounced for machine malfunctions than for human errors. This contradicts previous findings that reported behavioral and electrophysiological responses to errors being larger when they are committed by a human agent than if they are caused by machine malfunctions. Our results might suggest that automated systems are expected to operate predictably and, as a consequence, in interactive settings, errors committed by such systems are more salient and elicit a larger prediction error signal than if the same mistake is made by a human agent.

1. Introduction

When performing various tasks, human agents often make use of outside help: They cooperate with other people and they relegate parts of the task to devices designed to support their actions. Beside the obvious benefits, this comes with the cost that agents are required to monitor partner mistakes and device malfunctions. Previous research indicates that observed errors are processed in a similar manner as if one had committed the error themselves (Bates et al., 2005; De Bruijn et al., 2012; Miltner et al., 2004; Schuch & Tipper, 2007). But the question remains whether this is only true if the error is caused by another human agent, or it also applies to device malfunctions. While there have been a few studies investigating this question (Desmet et al., 2014; Huberth et al., 2019, Somon et al., 2019), they mostly focused on supervisory settings, where participants only observed another agent's actions. (Or in cases where the observer also performed a task, the two agents' actions were independent of each other.) The relevance of the partner's errors, however, is arguably higher in interactive tasks where the two agents' actions are connected. In the current study we used a leader-follower setting to investigate whether there are differences in the processing of observed errors, as reflected in error-related event-related potential (ERP) components, when they are committed by a human co-actor or by a machine.

Previous studies on error observation have indicated that phenomena that have been found in relation to one's own errors also manifest when somebody sees another agent committing an error (for a review, se Musco et al., 2023). This applies both to behavioral effects (de Bruijn et al., 2012; Pfister et al., 2020; Wang et al., 2015, 2016)—like post-error slowing (PES)—and to error-related components in neurophysiological data (Bates et al., 2005; de Bruijn & von Rhein, 2012; van Schie et al., 2004)—like the error-related negativity (ERN) or the error positivity (P_E). Most of the work on error observation focuses on

* Corresponding author.
** Corresponding author at: Department of Psychology, Trier University, Johanniterufer 15, Trier 54290, Germany.
E-mail addresses: bence.neszmelyi@gmail.com (B. Neszmélyi), roland.pfister@uni-trier.de (R. Pfister).

https://doi.org/10.1016/j.biopsycho.2024.108965
Received 30 August 2024; Received in revised form 9 December 2024; Accepted 12 December 2024
Available online 13 December 2024

monitoring human agents, but error-related neurophysiological and behavioral responses following machine malfunctions have also been examined (Castellar et al., 2011; Desmet et al., 2014; Somon et al., 2019). With increasing reliance on technological advancements, the relevance of such investigations could substantially increase in the near future. Comparing the processing of human and machine errors could potentially reveal the fundamental mechanisms underlying social error monitoring: If the processing of observed errors is based on matching the agent's movements onto the observer's own motor repertoire (Bates et al., 2005; van Schie et al., 2004), or on aligning the agent's and the observer's goals (Castellar et al., 2011; Koban et al. 2012), the processing of human errors should have a distinguished status. However, if errors are only interpreted in relation to the observer's goals (de Bruijn et al., 2012; Castellar et al., 2011), it might not make a difference whether the error was committed by a human or by an artificial agent. Research on this topic has yielded ambiguous results. A study focusing on observation-related post-error slowing (oPES[1]; Castellar et al., 2011) did not find a difference in oPES after human and machine errors. Studies using neurophysiological measures also found mostly similar patterns in error-related fMRI (Desmet et al., 2014) and EEG signals (Huberth et al., 2019, Somon et al., 2019) when monitoring human and artificial co-actors, although two of the studies also reported differences: Desmet et al. (2014) observed activation specific for the processing of human errors in the dorsal premotor cortex, right superior parietal lobule, left postcentral gyrus, and temporal gyrus, while Somon et al. 2019 found a difference between ERPs elicited by human and machine errors in the fronto-centro-parietal regions 400–500 ms after the erroneous response was observed.

The studies mentioned in the previous paragraph investigated the social aspect of error monitoring in a purely supervisory setting, where the observers did not perform a task themselves, or if they did, the two agents' actions were independent of each other. Many real-life scenarios, by contrast, come with strong interdependence of the co-actors' tasks. One such type of interaction are the leader-follower scenarios. They come with clear hierarchical and temporal order of two agents' actions: The leader initiates the interaction and this serves as a cue for the follower's response. Thus, the follower's reaction can be interpreted as an action effect elicited by the leader's action. According to the sociomotor framework (Kunde et al., 2018; Neszmélyi et al., 2022; Pfister et al., 2024), such social action effects can play a similar role in action planning and action initiation as non-social effects that are elicited when interacting with the inanimate environment: When an associative link has been established between an action and its consequence (be it an inanimate sensory effect or a social effect), activating the representation of the effect (for example, by imagining or perceiving it) can in turn activate the action representation as well, leading to the automatic selection of the adequate motor plan and the initiation of the action. Compared to supervisory settings, such interactive tasks increase the relevance of the co-actor's actions for the observer: The fact that the co-actor's responses are social action effects from the leader's perspective—and as such, they play a central role in the selection and initiation of the leader's action and also provide feedback on it—increases the importance of these events for the leader. The increased relevance of the co-actor's actions could also impact how the observer (i.e., the leader) processes the follower's errors.

In line with the above reasoning, a study by Weller and colleagues (2018) showed increased oPES as well as increased observed-error positivity (oP_E) in an interactive, leader-follower situation compared to the simple supervision of another agent. If social performance monitoring is indeed enhanced in the leader-follower setting, this task type might be also well suited to examine differences in interactions with machines and humans. Beside the usual factors that distinguish cooperation with another human from interactions with various devices (e.g., processing of biological movement: Thompson et al., 2005; unique effector systems: Crivelli & Fridlund, 2018; Neumann et al., 2014), there might be additional human-machine differences specific to action effect processing: Different response tendencies are associated with humans and machines, and this can also influence both the expectations regarding the elicited effects as well as the evaluation of expected versus actual outcome (Bigelow, 1998; Dzindolet et al., 2002, 2003; Striano et al., 2005). While machines are generally considered consistent and predictable, people grow accustomed to imperfect contingencies from an early age when interacting with other human agents. Thus, compared to interactions with machines, when having another human agent in the follower role, leaders might expect a higher error rate, might form less precise predictions regarding the expected responses, and might be less stringent in evaluating deviations from the predicted outcome. Following a similar line of thought, Pfister and colleagues used a leader-follower setting to compare error monitoring in interaction with other humans and with machines (Pfister et al., 2020). They found that compared to human-machine interactions, oPES is more pronounced when interacting with a human agent. These results were the first to indicate specific behavioral responses to human and machine errors and they suggest that the processing of observed errors is enhanced in social interactions—to some degree contradicting the expectation that machine errors would be more salient and as such they could have a larger impact.

The goal of the current study was to adjust the experimental design used by Pfister et al. (2020) to accommodate the assessment of error-related ERP components and to use this approach to investigate if larger behavioral responses to human errors (compared to machine malfunctions) are also reflected in enhanced amplitudes of the waveforms associated with error observation. We focused on ERP components that are well established in research on self-committed errors and in research on error observation as well, and that are thought to reflect different stages of error processing: The error-related negativity (ERN; Falkenstein et al., 2000; Gehring et al., 1993) is a frontocentral negative deflection—peaking ca. 80 ms after the action—that has been associated with error detection and with mismatch between the actual response and the correct response (Coles et al., 2001). The error positivity (P_E) is a positive deflection—peaking between 100 and 200 ms after the incorrect response—with a centro-parietal distribution. This component has been suggested to reflect error awareness (Nieuwenhuis et al., 2001, Overbeek et al., 2005), emotional response to the error or the adjustment of the response strategy (Falkenstein et al., 2000). Importantly, similar components have also been recorded in experimental contexts where participants did not commit the error themselves but only observed another agent executing an erroneous action or an action with a negative feedback[2] (Bates et al., 2005; Kang et al., 2010; Miltner et al., 2004; Van Schie et al., 2004). In the current study, we used the

[1] We use the following abbreviations for ERP components and behavioral effects that were investigated in the study: oPES = observation-related post-error slowing; oERN = observed-error-related negativity; oP_E = observed-error positivity.

[2] Despite the similarities, there are also differences between the components elicited by self-committed and observed errors (Koban et al. 2010; Picton et al., 2012).

observational versions of the error-related ERP components (i.e., observed-error-related negativity: oERN[3]; observed-error positivity: oP_E) to compare the processing of human and machine errors. We hypothesized that in a leader-follower setting, these components will be more pronounced in the social version of the task where the errors are related to a human co-actor's response to the leader's action compared to the non-social version where the errors are induced by machine malfunctions. Given the debate regarding the specific error-processing steps reflected in the oERN and oP_E (e.g., Falkenstein et al., 2000; Holroyd & Coles, 2002), we did not formulate strong predictions about which component would be more strongly influenced by the co-actor type. However, Weller et al. (2018) suggested that, compared to a passive observational setting, the mechanisms specifically involved in monitoring the social effects of one's own actions are predominantly reflected in the oP_E. Accordingly, we hypothesized that the social vs. non-social distinction would manifest more prominently in this ERP component. This hypothesis aligns with the proposition that the (o)P_E is associated with the conscious evaluation of errors, whereas the (o)ERN reflects more automatic error detection processes (Nieuwenhuis et al., 2001; Overbeek et al., 2005). Given that the social and non-social interactions were characterized by identical physical stimulation, we anticipated that the top-down construal of the task as either social or non-social would exert a greater influence on the ERPs associated with later, evaluative processing stages.

2. Method

2.1. Participants

Previous studies indicated that both oERN and oP_E were robust effects. Power analysis suggested that a sample size of six participants would be required to replicate the oERN effect obtained with an experimental method similar to the current one by Weller and colleagues (2018) and a sample size of seven for the replication of the oP_E effect reported in the same study. (Both calculated with an alpha level of.05 and a power of .80.) There were, however, no results available to make an educated guess about the impact of co-actor identity on the two ERP components. We recruited a sample of twenty-five participants. Students of the University of Würzburg participated in the experiment for monetary compensation. Informed consent was obtained from all participants. One participant was excluded, because after the application of the behavioral and EEG-based rejection criteria, the number of remaining trials was lower than the preregistered critical value. A second participant was excluded due to technical problems during EEG recording. The final sample consisted of twenty-three participants (mean age = 24.27; age range: 18 – 32; left-handed: 1, male: 4; one participant's demographic data were lost, so the description of the sample is based on 22 participants). At an alpha level of.05 and with a power of.80, this sample size can reveal an effect size[4] of the magnitude $dz = 0.61$.

Fig. 1. The experimental setup. The leader (participant) sat in a sound-proof EEG chamber while the follower (experimenter) sat outside the chamber. The figure shows how the feedback stage in the social task was explained in the instruction: The final visual effect on the leader's screen is elicited by the follower's keypress. (In reality, the follower's action was not required for the final effect to appear).

2.2. Apparatus and stimuli

The experiment took place in a laboratory with an electrically shielded, sound-proof cubicle and a workspace outside the cubicle (Fig. 1). The participant was seated inside the cubicle, while the experimenter, who was also the participant's interaction partner in the social task, sat outside. E-Prime 3.0 software (Psychology Software Tools, 2016) was used for stimulus presentation and for recording responses. The cubicle and the workspace for the experimenter were both outfitted with a 17" monitor (75-Hz refresh rate) and a standard QWERTZ keyboard. Target stimuli were displayed in white against a black background. Simple geometric shapes were presented as targets for the participant (circle, triangle, rectangle and hexagon; size: 2 cm × 2 cm). Targets appearing on the experimenter's screen were four different, complex, snowflake-like shapes. Participants responded with the C, V, N and M keys, whereas the experimenter used the R, T, U and I keys. The keys were operated with the index and middle fingers of each hand. All response keys were marked with color labels. Corresponding to the layout of the response keys, four blue rectangles (size: 3.2 cm × 3.2 cm) were also displayed on the participants' screen (but not on the experimenter's) in a horizontal row, below the midline.

[3] It is debatable whether the error-related negative peak that we assessed in the current study should be categorized as feedback-related (oFRN) or as error-related negativity (oERN). The stimuli eliciting the ERPs that we examined can be regarded as feedback signaling whether the co-actor's response (or the interaction as a whole) was correct. However, the visual effects in the present study do not simply signal if the action was correct, they also represent the co-actor's actions: They provide information both on the spatial (which key was pressed) and on the temporal aspect of the action (when was the key pressed). Thus, perceiving the action effects generated by the co-actor is in many aspects similar to observing their actions directly. Conceptually similar phenomena have been labelled in past research as oERN and as oFRN as well (e.g., Huberth et al., 2019; Weller et al., 2018). We decided to use the former term because it emphasizes the monitoring of the co-actor's actions, and the choice to represent these actions by their visual effects was methodologically and not conceptually motivated. (This way it was easier to warrant that physical parameters of social and non-social action effects were identical.)

[4] To our knowledge there is no reliable formula available for sensitivity analysis in factorial within-subject designs, with considerable discrepancies between different approaches. Thus, we based the sensitivity power analysis on a paired *t*-test, since the interaction of the two-way repeated-measures ANOVA can be also assessed as a paired *t*-test on difference values. Although, for the ANOVAs we report partial eta-squared as the measure of effect size, the *dz* value obtained in the sensitivity analysis gives a good estimate of the magnitude of effect sizes that can be reliably revealed with the sample used in the current study.

2.3. Procedure

The experimental procedure was based on the method of Pfister et al. (2020), but the between-subject design was adapted to a within-subject setting, and we also made a few additional changes to accommodate the ERP technique. Participants' task was to perform keypresses in response to the targets presented on the screen before them (see Fig. 2). A response key was assigned to each of the four geometrical shapes (counterbalanced across participants), and participants were asked to press the corresponding key as fast as possible when a target stimulus was displayed on the screen (for 300 ms). When participants performed a correct response, with a random interval of 500–600 ms, the action was followed by one of the four squares on the participants' screen changing its color from blue to green for 400 ms. On 85 percent of the trials, the square that changed its color corresponded spatially to the key that the participant had pressed (*correct feedback trials*[5]). For example, if participants pressed the C key, the first rectangle from the left turned green. On the remaining 15 percent of the trials (*error feedback trials*), the color change affected one of the other squares (chosen randomly). The next trial started after an interval of 1200 ms. When participants pressed the wrong key in response to the geometrical shape, or if they pressed one of the keys before the presentation of the stimulus, the trial was interrupted, and a warning was presented on the screen. Although the timing and the location of the action effect (i.e., color change) was always controlled by the experimental software, participants were led to believe that the causal chain of events that resulted in the action effects depended on the experimental condition. They were informed that in some blocks the color change of the squares was controlled by the experimenter's actions, and thus, error trials signaled errors of the human partner. In other blocks, however, the instructions said that effects were controlled by the computer, and as such error trials were the results of machine malfunctions.

In the *social condition* participants were informed that their keypresses resulted in visual stimuli that were presented on the experimenter's screen (i.e., a complex, snowflake-like shape). The experimenter responded to each of the stimuli with a specific key. The mapping of the experimenter's responses to the shapes warranted spatial correspondence between the participant's keypress that elicited the stimulus and the experimenter's correct response to the same stimulus. Participants also received the information that the four squares on their screen showed the state of the experimenter's keys: When the experimenter pressed one of the keys, the corresponding square turned from blue to green. This meant that action effects (color changes) that were spatially compatible with the participants' actions signaled the experimenter's correct response while incompatible effects indicated an experimenter error. However, the information that participants received was not true. The timing and location of these effects were predetermined, and they were also displayed if the experimenter did not respond to the target presented on their screen.

In the *non-social condition,* participants were informed that the device of the experimenter had been disabled: Participants keypresses would not elicit stimuli on the experimenter's screen and the experimenter could also not affect the color change of the squares. Participants were told that the latter effects were elicited by the computer in response to participants' actions. The computer would change the color of the rectangle that is spatially compatible with the participant's action, but it would not always produce the correct response.

In addition to selecting the appropriate response to the target stimuli, participants were also asked to count the number of errors made by the co-actor (be it the experimenter or the computer program) and report it at the end of each block. To induce a larger variability in the number of errors, in each of the conditions there were 3 blocks with an error probability of .1, 3 blocks with an error probability of .2 and 1 block with an error probability of .15. (Thus, the overall percentage of errors was 15 percent in both conditions.) The order of the different percentage blocks was randomized across participants, but for a single participant the order was the same in both conditions.

A methodologically elegant property of the experimental design is that the two conditions only differed with regard to the participant's construal of the situation. However, this choice also resulted in the two conditions being perfectly identical from the participants' perspective, thus, we included some elements that could make the cover story believable: (1) When introducing the social task during the practice phase, participants were demonstrated that the setup indeed worked as described: The experimenter showed that based on the stimuli appearing on their screen, they could recognize which key had been pressed by the participant. The experimenter also announced a few of their own keypresses in advance, so participants could confirm that the visual stimuli on their screen correspond with the actions of the experimenter. (2) During the practice phase, participants switched places with the experimenter and had the opportunity to experience the task (both the social and the non-social condition) from their perspective. During this phase, the task functioned as it was described in the instructions: In the social condition, the final action effects on the leader's screen were indeed elicited by the follower's response. To demonstrate this, the experimenter who took the participant's place in the cubicle, gave verbal feedback on the participant's actions. In the non-social condition, participants experienced that the monitor and the keyboard outside the cubicle were disabled and they were also shown that in this phase, the computer generated feedback (color change of the rectangles) on the screen inside the cubicle appeared without the follower's contribution. (3) Before the start of each block, participants were asked to signal when they were ready to start the block by pressing the "c" key on the keyboard. A red rectangle was presented on the screen, which turned green when participants performed the keypress. In the social condition, a second rectangle was also presented on the screen and the instructions asked both the participant and the experimenter to perform a keypress to signal that they are ready to start. We assumed that this would strengthen the participants' impression that they performed the task alone (non-social condition) or in cooperation with the experimenter (social condition).

After participants arrived at the lab, they were presented with a consent form. Then, participants completed several practice phases with a step-by-step introduction of different elements of the task. First, participants received the instructions for either the social or for the non-social task (counterbalanced across participants) and it was demonstrated that the setup indeed worked as described in the cover story. Participants then completed a short version (16 trials) of the task that had been introduced. After that, the other task was introduced, and participants also completed a short version of that. In the final practice phase, participants were asked to switch places with the experimenter, and repeated the previous steps of the practice. After the practice phase, preparations for the EEG-recording were made. Participants then completed 14 blocks of 52 trials. They started with the condition that was presented to them first during the practice phase and completed seven blocks of this task version. They could pause shortly between blocks. Halfway through the experiment (after the 7th block has been completed), a longer break was scheduled, and after that, participants continued with the other condition, again completing seven experimental blocks. After finishing the reaction time task, participants filled in a questionnaire on the experiment. We asked them whether they believed the cover story about the interactive and individual tasks, and how they experienced the difference between the two conditions.

2.4. Behavioral measures

During the experiment we recorded the accuracy and the timing of participants' responses. Accuracy was recorded as a binary variable: If

[5] We use the term feedback to refer to observed events to distinguish these events from errors that were committed by the participants themselves.

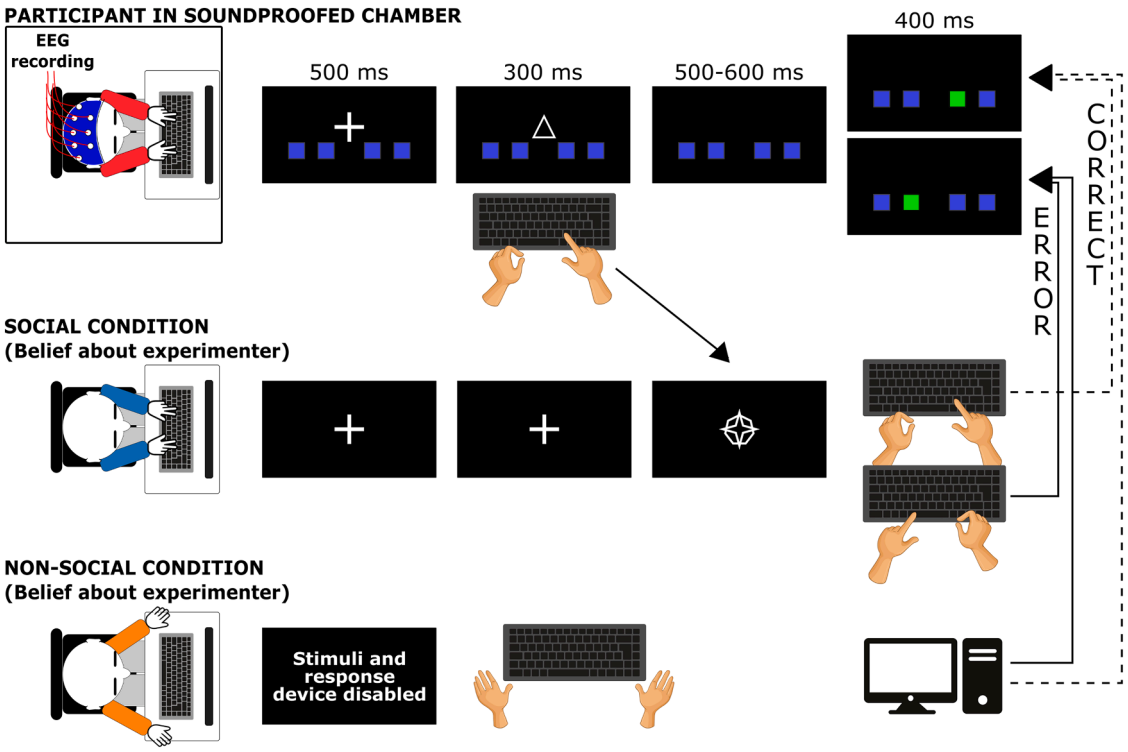

Fig. 2. The experimental method. The upper row shows participants' reaction time task and the visual effects that followed the participants' actions. The middle and bottom rows depict what the instructions told participants about the task of the co-actor (the experimenter). In the social condition, participants were led to believe that the visual effects displayed on their screen were elicited by their co-actor's correct or erroneous responses to their actions, while in the non-social condition they received the information that the co-actor does not contribute to the task, and the visual effects are generated by the computer.

participants pressed the correct key in response to the target stimulus, the trial was recorded as a hit, in all other cases (i.e., omission, miss, early or late response) it was recorded as an error trial.[6] Response times were measured as the interval between the presentation of the target and the response of the participant. Furthermore, we also assessed the accuracy of recalling the number of error trials at the end of each experimental block.

2.5. EEG recording and signal processing

A BrainVision QuickAmp amplifier with 32 active electrodes was used for recording the EEG signal from the participants. EEG was recorded with a 1000-Hz sampling rate using a 200-Hz online low-pass filter. Electrodes were positioned according to the 10–20 System, with additional electrodes placed at the mastoids. Horizontal and vertical electrooculogram was acquired from passive electrodes above and below the left eye (HEOG) and at the outer canthi of the eyes (VEOG). EEG was analyzed with the MNE Python package (Gramfort et al., 2014). The signal was re-referenced offline to the signal average of the mastoids and filtered using a 0.1-Hz high-pass filter and a 47.5–52.5-Hz band-stop filter. Eye movement artifacts were corrected by performing independent component analysis and rejecting components that correlated with signal at any of the EOG channels[7] (using an adaptive z-score-based threshold of 3.0). After that, the signal was re-filtered, using a 1—14-Hz band-pass filter for oERN analysis and a 0.1—14-Hz filter for oP_E analysis (Weller et al., 2018). 900-ms epochs, time-locked to the presentation of the feedback stimuli (color change of the squares), including a 100-ms pre-stimulus baseline, were extracted from the continuous EEG signal. For each error feedback trial, a matching correct feedback trial with identical effect location was selected, and only this subset was included in the analyses. EEG signal averages were calculated at FCz for oERN and at Pz for oP_E. For each participant and condition, the signal average of the feedback error and correct feedback trials was calculated. Then, using these individual averages, we calculated the *error feedback minus correct feedback* difference waveforms for each participant and condition. We determined the oERN time window as the 100-ms interval centered at the negative peak (between 200 and 350 ms after stimulus presentation) in the average of the grand average difference waveforms in the two conditions.[8] The oP_E time window was determined in a similar manner, using the positive peak (between 250 and 500 ms after stimulus presentation) in the signal average of grand average difference waveforms as the center of the 100-ms window. For each participant, condition and feedback accuracy, oERN and oP_E amplitudes were calculated as the signal average in the above time windows. Both amplitude values were calculated in relation to a −100–0-ms pre-stimulus baseline.

[6] The analysis of accuracy data was also performed by only considering misses as error trials, but since the same pattern of results was observed as with the wider range of errors, we only report the results obtained with the latter approach.

[7] This resulted in the exclusion of $M = 1.73$, $SD = 0.70$, 95 % CI [1.42, 2.04] components on average.

[8] Initially we planned to determine the time windows for amplitude measurement using the average (and not the difference) of the ERPs related to correct feedback and error feedback trials. This plan was also reported in the preregistration. However, since oERN and oP_E peaks were only apparent in the error feedback trials, we used the difference waveforms instead, while retaining other aspects of the planned analysis (electrodes and time ranges of interest). Using this method for determining the time windows for amplitude measurements can bias the error feedback vs. correct feedback comparison. Thus, although we also report the feedback main effect for the two-ways ANOVAs, we limit the interpretation to the condition × feedback interaction.

2.6. Statistical analyses

We planned to exclude participants from analyses if they were off by ±2 trials when reporting the number of error feedback trials in at least 8 blocks, but all participants fulfilled this requirement. From the analysis of response accuracy, only trials were excluded where participants themselves made a response error on the preceding trial (action started before the presentation of the target, no response in the 1500-ms time window after target presentation, or wrong response to the target). When analyzing reaction times (i.e., post-error slowing), beside response errors on the preceding trial, response errors on the actual trial and a response time of 2.5 *SD* over or below the cell average (for each participant, condition and previous feedback) were also applied as additional rejection criteria. For EEG analysis, rejection criteria applied for the reaction-time analyses (response errors on the preceding or on the actual trial, reaction times 2.5 *SD* over or below the cell average) were complemented by the exclusion of epochs where the signal range exceeded 150 μV at any channel. Participants for whom less than 25 trials remained in either of the cells, were not included in statistical analyses. Trial numbers considered for various analyses are summarized in Table 1.

Due to the experimental design, the number of error feedback trials was substantially lower than the number of correct feedback trials. Since unequal sample sizes can bias the estimation of the effects, for all three major analyses (post-error accuracy, post-error reaction times and error-observation-related ERP components), from each participant's full set of correct feedback (and corresponding post-correct-feedback) trials we randomly selected a subset that matched the number of error-feedback/post-error-feedback trials available for that participant. The only restriction during the selection of the subsample was that matching error-feedback and correct-feedback trials would be characterized by identical location of the feedback stimulus (i.e., color change of the same rectangle).

The absolute difference between the reported and actual number of error feedback trials was calculated for each participant and block and averaged across blocks within a condition. The recall-accuracy values obtained in the social and non-social conditions were then compared with a paired *t*-test. Average reaction time and the percentage of response errors was calculated for each participant, condition and previous feedback (correct or error) and both measures were analyzed with Condition × Previous feedback repeated-measures ANOVAs.

We estimated the oERN and oP_E amplitudes for each participant, condition and feedback (correct feedback, error feedback). Then the two dependent variables were submitted to Condition (interactive, individual) × Feedback (correct, error) repeated-measures ANOVAs. To assess electrode effects, we also calculated the ERP amplitudes at different electrodes (oERN at Fz, FCz and Cz and oP_E at Fz, Cz and Pz), and submitted the values to Condition × Feedback × Electrode repeated-measures ANOVAs. However, since the electrode factor did not influence the effect of interest for either oERN or oP_E (that is, no Condition × Feedback accuracy × Electrode interaction was observed) the results of these analyses are reported in the Supplementary Material (Figure S1 and Table S1). As an exploratory analysis, we also conducted a cluster-based permutation analysis to compare *error feedback minus correct feedback* difference waves in the social and non-social conditions. This analysis is also presented in the Supplementary Material (Figure S2). In all two-way ANOVAs (reaction time, accuracy, ERP amplitudes) the focus was on the Condition × Feedback or Condition × Previous feedback interaction. If needed, paired *t*-tests with Bonferroni correction were used for post-hoc analyses.

3. Results

3.1. Behavioral data

In the error counting task, participants generally performed well: On average, they reported a wrong number of error feedback trials in $M = 2.43$, $SD = 1.88$, 95% CI [1.62, 3.25] of 7 blocks in the non-social and $M = 2.17$, $SD = 1.83$, 95% CI [1.38, 2.96] of 7 blocks in the social condition. The average difference between the actual and the reported number of error feedback was $M = 0.57$, $SD = 0.71$, 95% CI [0.26, 0.87] in the non-social and $M = 0.46$, $SD = 0.44$, 95% CI [0.27, 0.65] in the social condition. The difference between conditions in the extent of counting errors was not significant, $t(22) = 0.74$, $p = .465$, $d_z = 0.15$.

Fig. 3 shows mean reaction times and accuracies for all experimental conditions. The analysis of reaction times indicated a main effect of previous feedback, $F(1, 22) = 12.29$, $p = .002$, $\eta_P^2 = .358$, which shows that actions after error feedback were initiated slower (non-social condition: $M = 675.54$, $SD = 102.11$, 95% CI [631.38, 719.69]; social condition: $M = 682.48$, $SD = 92.93$, 95% CI [642.30, 722.66]) than actions after correct feedback (non-social condition: $M = 648.34$, $SD = 91.86$, 95 % CI [608.61, 688.06]; social condition: $M = 669.85$, $SD = 107.47$, 95% CI [623.37, 716.32]). The main effect of condition, $F(1, 22) = 3.25$, $p = .085$, $\eta_P^2 = .129$,—with descriptively slower reaction times in the social condition compared to the non-social condition—was not significant, but the relatively large effect size suggests that replication with a larger sample could possibly reveal an effect. The same might be

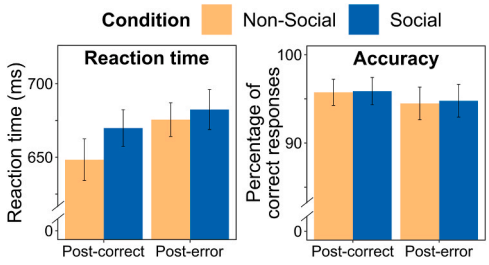

Fig. 3. Behavioral results. Group-average reaction time (left) and accuracy (right) on post-correct and post-error trials in the two conditions. Error bars indicate 95 % confidence intervals adjusted for the within-subjects design (Morey, 2008).

Table 1
The average number of trials (per participant) retained after applying exclusion criteria.

		Total	EEG-analysis		PES	PIA
Non-Social	Correct feedback	306.71±7.59	253.96±22.98 (47.48±8.69)*	Post-correct-feedback	255.61±21.36 (46.87±9.65)	275.04±13.40 (51.48±8.02)
	Error feedback	57.30±7.59	47.48±8.69	Post-error-feedback	46.87±9.65	51.48±8.02
Social	Correct feedback	308.61±6.87	254.65±22.21 (46.09±7.53)	Post-correct-feedback	257.17±20.15 (45.39±8.11)	276.70±12.89 (49.65±6.81)
	Error feedback	55.39±6.87	46.09±7.53	Post-error-feedback	45.39±8.11	49.65±6.81

*To have equal number of (post-)error-feedback and (post-)correct-feedback trials, in the case of the latter trial type a randomized selection procedure (that took the location of the final effect into consideration) was used for determining the trials included in the analyses. The final number retained after this selection procedure, which is equal to the number of (post-)error-feedback trials in the same condition, is shown in parentheses.

true for the Condition × Previous feedback interaction, $F(1, 22) = 2.31$, $p = .143$, $\eta_P^2 = .095$, although in this case, the effect size is somewhat smaller. Even if we believe that the numerically larger post-error slowing in the non-social condition reflects a real effect, this would be the opposite of the effect that was reported by Pfister et al. (2020), who reported increased post-error slowing when monitoring human errors.

The percentage of correct responses was not influenced by previous feedback (correct vs. error), $F(1, 22) = 0.16$, $p = .689$, $\eta_P^2 = .007$, nor by condition (social vs. non-social), $F(1, 22) = 1.37$, $p = .255$, $\eta_P^2 = .059$, nor did the two factors interact, $F(1, 22) = 0.28$, $p = .600$, $\eta_P^2 = .013$.

3.2. Error-related ERPs

Fig. 4 shows grand-average ERPs for the main electrodes of interests, accompanied by their topographic distribution. Corresponding mean amplitude statistics are summarized in Fig. 5.

At FCz, oERN peaked 238 ms after the onset of the feedback stimulus, and individual amplitude values were calculated as the signal average in the 100-ms time window centered at this point. The analysis at FCz indicated a significant main effect of feedback (correct vs. error), $F(1, 22) = 6.81$, $p = .016$, $\eta_P^2 = .236$, and a significant Condition x Feedback interaction, $F(1, 22) = 6.42$, $p = .019$, $\eta_P^2 = .226$. The condition main effect was not significant, $F(1, 22) = 0.89$, $p = .355$, $\eta_P^2 = .039$. These results indicate that amplitudes related to error feedback were more negative than amplitudes related to correct feedback and that this difference was larger in the non-social than in the social condition (Figs. 4 and 5). This was supported by post-hoc *t*-tests with Bonferroni correction: In the non-social condition the difference between amplitudes in correct feedback and error feedback trials was significant, $t(22) = 3.34$, $p_B = .006$, $d_z = 0.70$. In the social condition a significant difference was not observed, $t(22) = 1.55$, $p_B = .269$, $d_z = 0.32$.

At Pz, oP_E peaked 429 ms after the onset of the feedback stimulus, and the 100-ms time window centered at this point was used for the amplitude assessment. The analysis at Pz indicated a significant main effect of condition, $F(1, 22) = 6.50$, $p = .018$, $\eta_P^2 = .228$, as well as a significant main effect of feedback, $F(1, 22) = 173.96$, $p < .001$, $\eta_P^2 = .888$. The Condition × Feedback interaction was not significant, $F(1, 22) = 0.60$, $p = .446$, $\eta_P^2 = .027$. These results suggest overall more positive oP_E amplitudes in the non-social compared to the social condition, and more positive amplitudes for error feedback than for correct feedback.

In the post-experiment questionnaire, participants indicated an average belief level of $M = 57.73$ (on a 0–100 scale), $SD = 28.31$, 95% CI [45.17, 70.28] in the cover story (i.e., that they indeed interacted with the experimenter and with the computer in the two conditions). As an exploration of the influence of belief in the experimental manipulation, we performed the ANOVAs with observed-error-related ERP components and the ANOVAs with accuracy and reaction time with belief in the experimental manipulation as an additional categorical predictor: Twelve participants reported a belief of at least 50 percent. They were categorized as “believers” while the remaining part of the sample was categorized as “sceptics”. The belief factor did not have a significant impact on any of the main effects or interactions. (See the Supplementary Material for more details.)

4. Discussion

We examined participants’ electrophysiological responses to errors committed by their partner when interacting with another human agent and with a machine. Based on previous results (Pfister et al. 2020), we expected that error-related ERP components would be more pronounced when errors are committed by a human co-actor. Instead, in the oERN time window, we observed an enhanced error-related signal component (i.e., the difference between ERPs induced by error feedback and by correct feedback) when the errors were generated by the computer, while there was no evidence for an impact of co-actor identity in the oP_E time window. Although ERPs were the main focus of our study, we also examined behavioral responses related to observed errors. We did not find any impact of error observation on accuracy, but similarly to findings reported in previous studies (Castellar et al., 2011; Pfister et al., 2020; Sacheli et al., 2021; Weller et al., 2018), reaction time increased on trials following errors committed by the co-actor. In contrast with the results of Pfister et al. (2020), however, this oPES effect was not enhanced when participants were cooperating with another human agent. If anything, oPES seemed slightly more pronounced during human-machine cooperation. This would be consistent with the oERN results, but the difference between conditions was not significant and since the effect seemed relatively small, a replication with a larger sample would be necessary to assess whether oPES depends on the type of agent who is committing the error. In summary we can conclude that while past research found either no difference between the processing of human and machine errors, or an enhanced error processing when the error is committed by a human agent (Castellar et al., 2011; Huberth et al., 2019; Pfister et al., 2020; Somon et al., 2019), our study rather indicated stronger error-related effects in relation to machine errors. As we discuss in the following, a reason for this deviation from previous results might be that the relevance of the observed agent’s actions for the observer is different in a supervisory context and in a leader-follower setting.

4.1. Error-related ERPs

The latency and the topographical distribution of the error-related components obtained in the current study correspond to previous findings (see Weller et al., 2018, for example), thus, it is likely that these components reflect the same error-processing mechanisms that have been associated with (o)ERN and (o)P_E in prior research. What exactly these mechanisms are is a topic of longstanding debate (e.g., Falkenstein et al., 2000; Holroyd & Coles, 2002), but a relatively consistent finding across studies is that only P_E amplitudes are influenced by the conscious detection of errors, while the ERN seems independent of error awareness (Nieuwenhuis et al., 2001; Overbeek et al., 2005). Thus, it seems plausible that ERN reflects an initial, automatic stage of error monitoring, while conscious, evaluative processes—that might also involve the preparation of adjustments made in response to the errors—have a stronger connection with the P_E component (Falkenstein et al., 2000; Nieuwenhuis et al., 2001). Although in general, the error-related ERP components obtained in the current study can be regarded as typical, the pattern that they showed in the two experimental conditions contradicted our expectations in multiple aspects.

Because the two conditions were identical with regards to physical stimulation, we expected that if an impact of the co-actor identity would be observed, it would more likely be present at the later, evaluative processing stage. That would have been also consistent with the findings of Somon et al. (2019), who found a difference between the processing of human and machine errors in the 400–500 ms range (after observing the error)—although in more anterior locations compared to where we assessed the oP_E. Furthermore, when comparing error processing for commanded and observed actions, while keeping the physical parameters of the conditions constant, Weller et al. (2018) also only found a difference in oP_E amplitudes (but not in the oERN component), which could suggest that beliefs regarding the cause of the errors only influence later processing stages. In the present study, however, the difference in the processing of human and machine errors emerged in the oERN time window and not in the oP_E time window. Regardless if the oERN recorded in our study reflects mismatch between representations of the correct and the actual response (Falkenstein et al., 1991; Hewig et al., 2011), the comparison process itself (Falkenstein et al., 2000) or the discrepancy between expected and actual outcome (Holroyd and Coles, 2002; Nieuwenhuis et al., 2004), this finding suggests that belief regarding the co-actor’s identity influences relatively early stages of

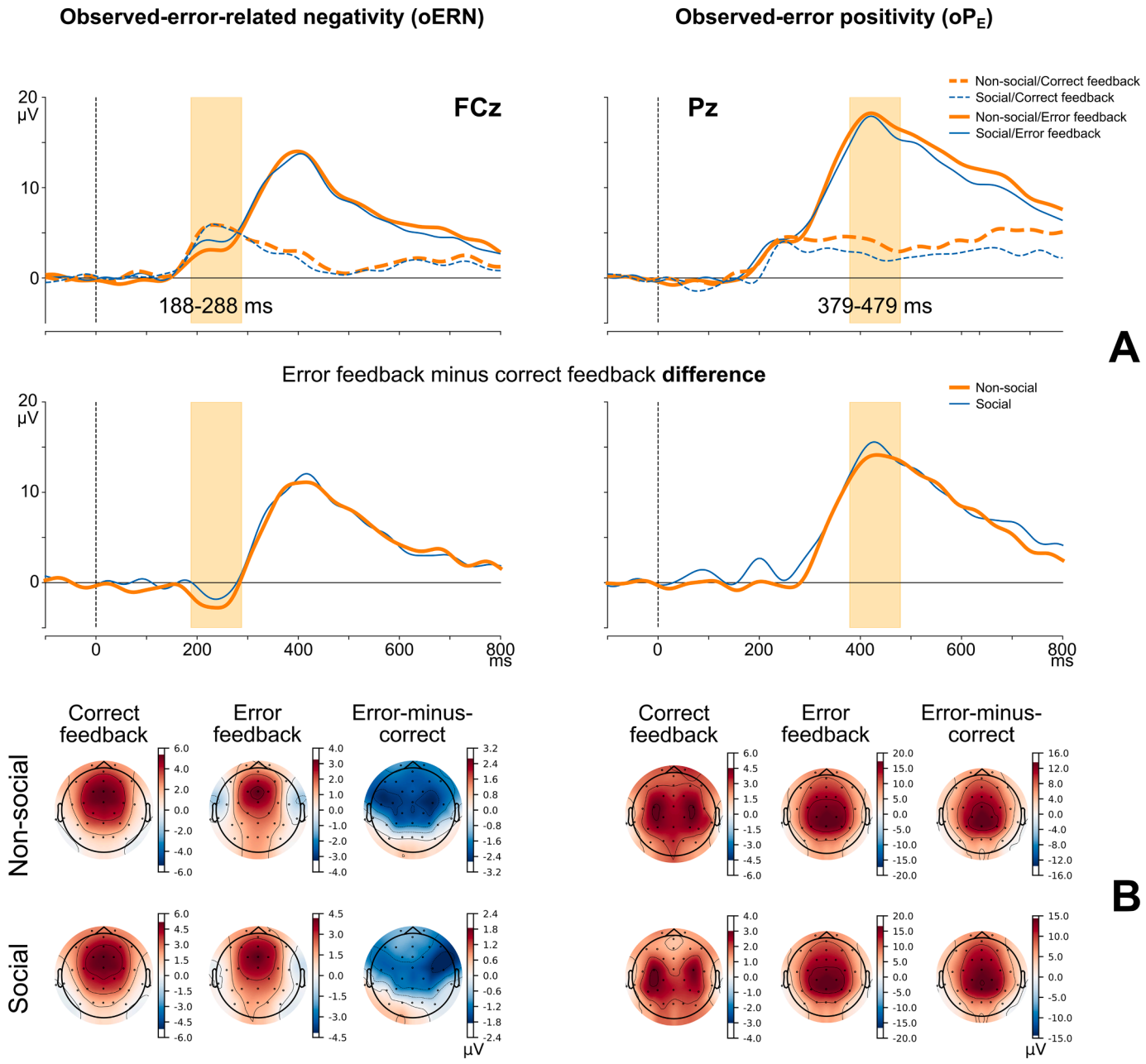

Fig. 4. ERPs in the social and non-social condition. The upper row in Panel A shows the ERPs obtained from correct feedback and error feedback trials, while the lower row shows the difference waveforms calculated from these ERPs (error feedback minus correct feedback). Panel B shows the topographical distribution for the two ERP components of interest, averaged in the 100-ms time windows marked by the orange bars in Panel A. Topographical maps are presented for all four combinations of condition and feedback (per component) as well as for the error-minus-correct difference waveform.

error processing. In the oP_E time window we only found a main effect of condition, indicating that regardless if the feedback was correct or incorrect, the late positive component was larger for non-social than for social effects. This might show that feedback in general was regarded as more relevant when interacting with a machine. A possible explanation for this is that consistent reactions from automated systems could lead to machine responses being interpreted as reliable feedback on one's own actions (i.e., concealing the indirect nature of the action effects), while in the case of human interaction there are more degrees of freedom associated with the action effects and as such the information content of these stimuli might be regarded as less valuable.

Beside the time window where the impact of manipulating participants' belief about the co-actor's identity was observed, the direction of the effect was also contrary to our expectations. Previous studies suggested that error monitoring is enhanced when observing human as compared to machine errors (Pfister et al., 2020; Somon et al., 2019). In apparent contrast with these studies, we observed larger oERN amplitudes in the non-social condition. However, our findings might be reconciled with prior knowledge about human-human and human-machine interactions: According to Lewandowsky et al. (2000), when it comes to task allocation, humans trust automated systems more than they trust other human agents. On the one hand, this can have the effect of reduced attention and increased complacency (i.e., operators becoming too trusting of the supervised system and not recognizing errors as a consequence: Moray et al., 2000; Sheridan & Parasuraman, 2006). On the other hand, the expectation of consistent and reliable functioning can also make errors made by machines more salient. This stands in contrast with human agents, who can be expected to make

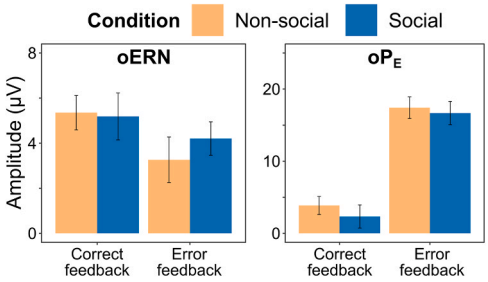

Fig. 5. Group-average ERP amplitudes in correct feedback and error feedback trials in *the two conditions*. Amplitude values were calculated in the oERN (left) and oP_E (right) time windows. Error bars indicate 95 % confidence intervals adjusted for the within-subjects design (Morey, 2008).

mistakes—as reflected in the saying "to err is human" (Pope, 1711). Furthermore, even disregarding involuntary mistakes, the action-effect contingency is arguably reduced when the effects are the reactions of a human co-actor as opposed to machine operations (Kunde et al., 2018; Watson, 1979): Humans are perceived as having the free choice to select from a wide range of available responses while machines operate in a more deterministic manner. Increased error probability and reduced action-effect contingency could both contribute to participants forming less accurate predictions regarding the outcome of the actions or putting less weight on such predictions. As a result, when producing social effects, the interpretation of errors might be less stringent and the prediction error signal induced by an erroneous outcome less pronounced—which could be reflected in the (o)ERN (Yasuda et al., 2004).

While our study was not designed to contrast different explanations of error-related ERPs, the present findings align well with expectancy-based accounts of the ERN/FRN (Holroyd & Coles, 2002; Alexander & Brown, 2011). These accounts suggest that the ERN and FRN do not rely on a cognitive mechanism specifically dedicated to recognizing errors or response conflict; rather, they reflect expectancy violations. For example, the reinforcement-learning theory of the FRN/ERN (Holroyd & Coles, 2002; Holroyd et al., 2005) posits that these components represent negative deviations from expected outcomes. Recent approaches, such as the prediction of response outcome theory (Alexander & Brown, 2011), go a step further by proposing that these components can be elicited by prediction violations regardless of outcome valence—that is, even by outcomes that are more positive than expected.

In this context, the increased oERN for machine errors observed in the present study may reflect a larger prediction error, as participants likely formed more specific expectations about human-machine interactions while demonstrating greater tolerance for diverse action outcomes in human-human interactions. Studies supporting expectancy-based explanations often establish predictions by manipulating the frequency of different event types (e.g., Ferdinand et al., 2012; Wessel et al., 2012). This applies to our study as well. However, in our experimental setup, error probability and action-effect predictability were identical across the social and non-social conditions. Thus, if participants perceived differences between these conditions, such perceptions were likely shaped by prior experiences with human and artificial agents or by beliefs and attitudes stemming from those experiences.

Krigolson and Holroyd (2006, 2007) also raised the possibility that ERN and P_E do not reflect different stages of error processing, but different error types. According to this idea, ERN is induced by high-level (i.e., failure to reach goal) and P_E by low-level errors (e.g., discrepancy between desired and actual motor command). The fact that in the present study, the social and non-social conditions were identical on the physical level could explain why there is no difference in the low-level goals and as such in oP_E. It is possible, however, that high-level goals are different in human-human and human-machine interactions. Due to the deterministic nature of machines, the goal of a human agent when operating a device could be to elicit a specific effect/response, while agents might have less specific goals in mind when cooperating with a human an agent (e.g., to elicit any response). As a consequence, the same incompatible action effect might be regarded as large deviation from the goal in the non-social condition and only as a small error (or even no error) in the social condition, which could explain the timing and location (oERN instead of oP_E) as well as the direction of the observed effect.

The above explanations are based on the idea that larger (o)ERN amplitudes reflect enhanced error monitoring (Hajcak et al., 2005). A study by Carp et al. (2009) indicated that this presumption might need revision in certain cases. The authors reported inconsistencies with regard to the correlation between error-related ERP amplitudes and observer-agent similarity: similarity and oP_E showed a positive correlation, whereas similarity and oERN exhibited a negative correlation. Since the authors assessed error-related components in the correct-minus-error difference waveforms, they explained the unexpected phenomenon by higher similarity resulting in increased negative, ERN-like peaks for the observation of correct actions, and as such they assumed that it is actually increased monitoring that resulted in the smaller oERN amplitudes. Although this interpretation does not apply to our results, it demonstrates that smaller oERNs do not necessarily indicate reduced or less effective monitoring. Furthermore, the oERN being smaller in the social setting of the present study is also consistent with the results of Carp et al., since the similarity between the participants (the observer) and a human co-actor is larger than the human-computer similarity. The picture seems, however, more complicated, if we also consider a study of Kang et al. (2010), which reported a positive correlation between oFRN and closeness between observer and actor.

Interestingly, while a significant oERN peak was reported by Weller et al. (2018) in a leader-follower setting with a second participant in the follower role (i.e, very similar to the social condition of the present experiment), we only found a significant difference in the oERN time window between amplitudes related to correct feedback and error feedback when the feedback was generated by the computer and not when it was produced by the actions of a human co-actor. On the one hand, the lack of an effect in the social condition might be in part a power issue, since a small, non-significant difference ($d_z = 0.30$) was also present in the current study. On the other hand, the ERP assessment method that was used in both studies (defining analysis time windows via signal peaks in the error-minus-correct difference waveforms) is only adequate for estimating between-condition differences in the error-related effects and not for an accurate assessment of the magnitude of the effect itself (i.e., the difference between amplitudes recorded in error and correct trials).

4.2. Behavioral results

The oPES observed after the co-actor's errors in similar leader-follower settings (Pfister et al., 2020; Weller et al., 2018) was also evident in the current study. However, in contrast with Pfister et al. (2020) we did not find increased post-error slowing in the social condition. If anything, post-error slowing seemed slightly larger in the non-social condition (although the difference between conditions was not significant), which would be in line with the oERN effect that was observed. Although the experimental design is based on the method used by Pfister et al., with a few adjustments that enable the assessment of ERPs, there is a key difference that might explain inconsistencies in the oPES results: Belief regarding the co-actor's identity was manipulated in a within-subjects manner in the present study, whereas Pfister et al. used a between-subjects design. In the within-subjects setting, it is more difficult to prevent participants from becoming suspicious with regard to the cover story. Efficiency of the belief manipulation or participants' search for evidence that would confirm or dispel their suspicions might account for differences in error-related behavioral patterns,

especially since—as discussed earlier—trust in automated systems can result in both decreased as well as increased attribution of resources to error monitoring. Furthermore, effect sizes for the Condition/setting (social, non-social) × Previous feedback (correct, error) interaction were relatively small in both experiments (η_P^2 =.07 in Pfister et al., 2020 and η_P^2 =.09 in the present study[9]), thus a replication with a larger sample would be required to obtain reliable evidence regarding the impact of co-actor identity on oPES.

4.3. Limitations and future directions

Our experimental paradigm was built on the assumption that we could induce different beliefs regarding the cause of the errors in the social and in the non-social condition. Results of the post-experiment questionnaire, however, indicated that the overall level of belief in the cover story was relatively low. Only slightly more than half of the sample reported a relatively high belief in the cause of the errors as explained in the instructions, while a substantial part of the sample can be interpreted as rather not believing the explanation they received. On the one hand, despite a large number of participants doubting the instructions, we found a significant effect, which could indicate that the difference between the processing of human and machine errors might be actually larger than the effect obtained in the present study. On the other hand, if the difference between conditions (in oERN and possibly also in oPES) is explained by the attribution of errors to different causes, then this effect should have been influenced by participants' belief in the cover story: For those who don't believe it, the two conditions should seem identical and in their case, there shouldn't be a difference between ERPs recorded in the social and non-social conditions. However, when including belief as a predictor in the ANOVAs of the electrophysiological and behavioral measures, this additional factor did not have an impact on the results (see Tables S2 and S3 in the Supplementary Material). A possible explanation for this might be that responses to the post-experiment questionnaire do not reflect participants' actual state of belief during the experiment: Participants responses might have been influenced by the aspiration to establish a positive image of themselves (i.e., expressing doubts regarding the cover story to appear more intelligent). It is also possible that participants who had a low belief score only started to be suspicious towards the end of the session, and the belief induction was still effective throughout a substantial part of the study. However, we also have to consider the possibility that the observed differences were not induced by participants' causal beliefs but by some other factor. It is possible, for example, that participants, being skeptical of the cover story, directed their attention toward finding evidence to confirm or disprove the information provided during the instructions. Depending on which aspects of the instructions they found suspicious or what type of evidence they sought, this could have led to a different allocation of cognitive resources across the two conditions, potentially contributing to the observed effects. However, the ANOVA that included belief as an additional predictor does not support this explanation. If the oERN difference between the social and non-social settings was linked to skeptical participants actively seeking evidence to confirm their doubts, the effect would have been expected to be more pronounced among those who did not believe the instructions. The analysis, however, revealed no such effects.

The post-hoc exploration of the influence of belief in the experimental manipulation lacked sufficient power to reliably detect smaller effect sizes. To ensure that the effect observed in the present study arises from differences in the monitoring of social versus non-social action effects—and not from skepticism toward the experimental manipulation—future research should aim to replicate these findings using modifications designed to enhance participants' belief in the instructions. Given the inherent transparency of within-subjects manipulations, achieving this might require adopting a between-subjects design, which would necessitate a substantially larger sample size.

In the present study we adopted the experimental design used by Weller et al. (2018) and by Pfister et al. (2020). In this design, error trials differ from correct ones not only in adherence of the co-actor to the task instructions, but also in action-effect compatibility (i.e., actions and visual effects are spatially compatible in correct trials but incompatible in error trials) and in frequency (correct trials are more frequent than error trials). Thus, it is possible that between-condition oERN differences are related to one of these factors and not to error processing. As a consequence, instead of referring to trials with alleged errors of the co-actor as error trials, categorizing them as oddball trials could have been more adequate (see Pfister et al., 2020). Despite these considerations, we retained the error-based terminology, because the fact that it was the oERN amplitude—which had been obtained previously in a wide range of error paradigms (see Musco et al., 2023)—that was affected by the identity of the co-actor, suggests a connection between error processing and the effect observed in the present study.[10] Furthermore, to enable a controlled comparison of social and non-social action effects, the leader-follower paradigm used in the present study focused specifically on the relationship between actions and their elicited effects. However, this approach necessarily overlooks other factors that characterize interpersonal interactions. Our findings suggest that performance monitoring may differ between social and non-social contexts. To further investigate the unique aspects of social error monitoring, future studies could employ experimental designs incorporating additional features of interpersonal interactions, such as observing the co-actor's actual actions (not only their effects), adopting interchangeable leader-follower roles, or allowing participants to correct each other's errors (see Neszmélyi et al., 2022).

In the current study, the instructions did not specify the reasons behind the errors made by the experimenter or the computer. During the practice phase, participants experienced the task from the follower's perspective, and the inherent difficulty of this part of the task provided an implicit justification for the experimenter's errors. However, participants might also have assumed that these errors were intentional. Similarly, machine errors could have been interpreted in various ways—as device malfunctions, intentional oddball events, or outcomes of randomization. While several studies have demonstrated that the alignment of observed errors with the observer's goals influences error processing (Castellar et al., 2011; de Bruijn et al., 2012; de Bruijn & von Rhein, 2012), to our knowledge, no studies suggest a similar influence based on the observer's assumptions about the alignment of the errors with the observed agent's goals. In other words, there is no evidence indicating that error processing differs depending on whether the observed agent intentionally chooses the wrong action (from the observer's perspective) or commits the error accidentally. Therefore, our interpretation of the current findings centers on the leader's perspective, positing that social action effects are generally less reliable or predictable than non-social ones. Nonetheless, the potential influence of intentionality and goal attribution to the co-actor should not be dismissed. Investigating these factors in relation to error processing represents an intriguing avenue for future research.

[9] This corresponds to $d_z = 0.31$ which is substantially smaller than the effect size that can be reliably shown with the present experimental design.

[10] However, it cannot be ruled out that this component represents a N200, which shares characteristics with the ERN/FRN. In conflict tasks, incongruent events are associated with an enhanced N200 (Gehring et al., 1992; Nieuwenhuis et al., 2003), and this component may have been influenced by spatial action-effect incompatibility in the present experiment. Within this framework, our findings could suggest that the spatial correspondence of causally connected events holds greater significance in interactions with machines, whereas interpersonal interactions allow for greater flexibility regarding the spatial relationship between agents' actions.

4.4. Conclusion

The current study provided strong evidence against our initial hypothesis that ERPs related to error observation would be more pronounced when they were induced by inadequate human responses to participant's actions as compared to erroneous computer responses. We even found some evidence for a reverse effect, with oERN amplitudes being larger in the non-social than in the social condition. A possible explanation for this unexpected effect is that machine malfunctions are associated with a larger prediction error than human errors. This could be based either on different goals (human: eliciting any response, machine: eliciting a specific response) or on different expectations (human: unpredictable response, machine: consistent response) when interacting with human agents and with machines in a leader-follower setting. Alternatively, the effect might be related to participants' suspicions regarding the cover story of the experiment and to their efforts invested in looking for evidence that would expose false information. To better understand differences between interacting with humans and with automated systems, future studies could further explore predictions based on these explanations.

Ethical approval

The present study was approved by the Ethics Committee of the Institute of Psychology at the University of Wuerzburg (GZ 2019–27), and the experiments were conducted in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments. Informed consent regarding participation in the study and publication of the collected data was obtained from all participants.

Open practices statement

Stimulus materials, raw data and scripts for the analyses are available on the Open Science Framework (https://osf.io/tzw8q). The preregistration can be viewed at https://aspredicted.org/BNW_W58.

Funding

This work was supported by the German Research Council (Deutsche Forschungsgemeinschaft; DFG) to R.P. (PF 853/2–2; PF 853/10–1).

CRediT authorship contribution statement

Roland Pfister: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Bence Neszmélyi:** Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors did not use generative AI or AI-assisted technologies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopsycho.2024.108965.

Data availability

A link to the repository where data and code are publicly available is shared in the Open practices statement.

References

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience, 14*(10), 1338–1344. https://doi.org/10.1038/nn.2921

Bates, A. T., Patel, T. P., & Liddle, P. F. (2005). External behavior monitoring mirrors internal behavior monitoring: Error-related negativity for observed errors. *Journal of Psychophysiology, 19*(4), 281–288. https://doi.org/10.1027/0269-8803.19.4.281

Bigelow, A. E. (1998). Infants' sensitivity to familiar imperfect contingencies in social interaction. *Infant Behavior Development, 21*, 149–162. https://doi.org/10.1016/S0163-6383(98)90060-1

Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., & Compton, R. J. (2009). Perceived similarity and neural mirroring: Evidence from vicarious error processing. *Social Neuroscience, 4*(1), 85–96. https://doi.org/10.1080/17470910802083167

Castellar, E. N., Notebaert, W., Van den Bossche, L., & Fias, W. (2011). How monitoring other's actions influences one's own performance: Post-error adjustments are influenced by the nature of the social interaction. *Experimental Psychology, 58*(6), 499–508. https://doi.org/10.1027/1618-3169/a000118

Coles, M. G., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology, 56*(3), 173–189. https://doi.org/10.1016/S0301-0511(01)00076-X

Crivelli, C., & Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends in Cognitive Sciences, 22*(5), 388–399. https://doi.org/10.1016/j.tics.2018.02.006

de Bruijn, E. R. A., & von Rhein, D. T. (2012). Is your error my concern? An event-related potential study on own and observed error detection in cooperation and competition. *Frontiers in Neuroscience, 6*. https://doi.org/10.3389/fnins.2012.00008

De Bruijn, E. R. A., Mars, R. B., Bekkering, H., & Coles, M. G. H. (2012). Your mistake is my mistake … or is it? Behavioural adjustments following own and observed actions in cooperative and competitive contexts. *Quarterly Journal of Experimental Psychology, 65*(2), 317–325. https://doi.org/10.1080/17470218.2010.545133

Desmet, C., Deschrijver, E., & Brass, M. (2014). How social is error observation? The neural mechanisms underlying the observation of human and machine errors. *Social Cognitive and Affective Neuroscience, 9*(4), 427–435. https://doi.org/10.1093/scan/nst002

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*(1), 79–94. https://doi.org/10.1518/0018720024494856

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*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology, 51*(2–3), 87–107. https://doi.org/10.1016/S0301-0511(99)00031-9

Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *The Journal of Neuroscience, 32*(35), 12087–12092. https://doi.org/10.1523/JNEUROSCI.1410-12.2012

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Probability effects on stimulus evaluation and response processes. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 198–216. https://doi.org/10.1037/0096-1523.18.1.198

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage, 86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology, 42*(2), 151–160. https://doi.org/10.1111/j.1469-8986.2005.00270.x

Hewig, J., Coles, M. G. H., Trippe, R. H., Hecht, H., & Miltner, W. H. R. (2011). Dissociation of Pe and ERN/Ne in the conscious recognition of an error: ERPs and consciousness. *Psychophysiology, 48*(10), 1390–1396. https://doi.org/10.1111/j.1469-8986.2011.01209.x

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*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General, 134*(2), 163–191. https://doi.org/10.1037/0096-3445.134.2.163

Huberth, M., Dauer, T., Nanou, C., Román, I., Gang, N., Reid, W., Wright, M., & Fujioka, T. (2019). Performance monitoring of self and other in a turn-taking piano

duet: A dual-EEG study. *Social Neuroscience, 14*(4), 449–461. https://doi.org/10.1080/17470919.2018.1492968
Kang, S. K., Hirsh, J. B., & Chasteen, A. L. (2010). Your mistakes are mine: Self-other overlap predicts neural response to observed errors. *Journal of Experimental Social Psychology, 46*(1), 229–232. https://doi.org/10.1016/j.jesp.2009.09.012
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*(3), 460–478. https://doi.org/10.3758/s13415-012-0091-0
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*(4), 360–374. https://doi.org/10.1080/17470911003651547
Krigolson, O. E., & Holroyd, C. B. (2006). Evidence for hierarchical error processing in the human brain. *Neuroscience, 137*(1), 13–17. https://doi.org/10.1016/j.neuroscience.2005.10.064
Krigolson, O. E., & Holroyd, C. B. (2007). Hierarchical error processing: Different errors, different systems. *Brain Research, 1155*, 70–80. https://doi.org/10.1016/j.brainres.2007.04.024
Kunde, W., Weller, L., & Pfister, R. (2018). Sociomotor action control. *Psychonomic Bulletin & Review, 25*(3), 917–931. https://doi.org/10.3758/s13423-017-1316-6
Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied, 6*(2), 104–123. https://doi.org/10.1037/1076-898X.6.2.104
Miltner, W. H. R., Brauer, J., Hecht, H., Trippe, R., & Coles, M. G. H. (2004). Parallel brain activity for self-generated and observed errors. *Errors, conflicts, and the brain. current opinions on performance monitoring* (pp. 124–129). Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences.
Moray, N., Inagaki, T., & Itoh, M. (2000). Situation adaptive automation, trust and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*(1), 44–58.
Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*, 61–64. https://doi.org/10.20982/tqmp.04.2.p061
Musco, M. A., Zazzera, E., Paulesu, E., & Sacheli, L. M. (2023). Error observation as a window on performance monitoring in social contexts? A systematic review. *Neuroscience Biobehavioral Reviews, 147*, Article 105077. https://doi.org/10.1016/j.neubiorev.2023.105077
Neszmélyi, B., Weller, L., Kunde, W., & Pfister, R. (2022). Social action effects: Representing predicted partner responses in social interactions. *Frontiers in Human Neuroscience, 16*, Article 837495. https://doi.org/10.3389/fnhum.2022.837495
Neumann, R., Lozo, L., & Kunde, W. (2014). Not all behaviors are controlled in the same way: Different mechanisms underlie manual and facial approach and avoidance responses. *Journal of Experimental Psychology: General, 143*(1), 1–8. https://doi.org/10.1037/a0032259
Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience Biobehavioral Reviews, 28*(4), 441–448. https://doi.org/10.1016/j.neubiorev.2004.05.003
Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38*(5), 752–760. https://doi.org/10.1111/1469-8986.3850752
Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, Behavioral Neuroscience, 3*(1), 17–26. https://doi.org/10.3758/CABN.3.1.17
Overbeek, T. J., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-a-vis the ERN/Ne. *Journal of Psychophysiology, 19*(4), 319–329. https://doi.org/10.1027/0269-8803.19.4.319
Pfister, R., Neszmélyi, B., & Kunde, W. (2004). Anticipated imitation. In O. Genschow, & E. Cracco (Eds.), *Automatic Imitation* (pp. 155–176). Springer. https://doi.org/10.1007/978-3-031-62634-0_8.
Pfister, R., Weller, L., & Kunde, W. (2020). When actions go awry: Monitoring partner errors and machine malfunctions. *Journal of Experimental Psychology: General, 149*(9), 1778–1787. https://doi.org/10.1037/xge0000748
Picton, L., Saunders, B., & Jentzsch, I. (2012). "I will fix only my own mistakes": An ERP study investigating error processing in a joint choice-RT task. *Neuropsychologia, 50*(5), 777–785. https://doi.org/10.1016/j.neuropsychologia.2012.01.011
Pope, A. (1711). *An essay on criticism*. Legare Street Press.
Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from ⟨https://support.pstnet.com/⟩.
Sacheli, L. M., Musco, M. A., Zazzera, E., & Paulesu, E. (2021). Mechanisms for mutual support in motor interactions. *Scientific Reports, 11*(1), 3060. https://doi.org/10.1038/s41598-021-82138-y
Schuch, S., & Tipper, S. P. (2007). On observing another person's actions: Influences of observed inhibition and errors. *Perception Psychophysics, 69*(5), 828–837. https://doi.org/10.3758/BF03193782
Sheridan, T., & Parasuraman, R. (2006). Human-automation interaction. In R. S. Nickerson (Ed.), *Reviews of human factors and ergonomics, 1* pp. 89–129). Santa Monica, CA: Human Factors and Ergonomics Society. https://doi.org/10.1518/155723405783703082.
Somon, B., Campagne, A., Delorme, A., & Berberian, B. (2019). Human or not human? Performance monitoring ERPs during human agent and machine supervision. *NeuroImage, 186*, 266–277. https://doi.org/10.1016/j.neuroimage.2018.11.013
Striano, T., Henning, A., & Stahl, D. (2005). Sensitivity to social contingencies between 1 and 3 months of age. *Developmental Science, 8*(6), 509–518. https://doi.org/10.1111/j.1467-7687.2005.00442.x
Thompson, J. C., Clarke, M., Stewart, T., & Puce, A. (2005). Configural processing of biological motion in human superior temporal sulcus. *Journal of Neuroscience, 25*(39), 9059–9066. https://doi.org/10.1523/JNEUROSCI.2129-05.2005
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*(5), 549–554. https://doi.org/10.1038/nn1239
Wang, L., Pan, W., Tan, J., Liu, C., & Chen, A. (2016). Slowing after observed error transfers across tasks. *PLOS ONE, 11*(3), Article e0149836. https://doi.org/10.1371/journal.pone.0149836
Wang, L., Tang, D., Zhao, Y., Hitchman, G., Wu, S., Tan, J., & Chen, A. (2015). Disentangling the impacts of outcome valence and outcome frequency on the post-error slowing. *Scientific Reports, 5*(1), 8708. https://doi.org/10.1038/srep08708
Watson, J. S. (1979). Perception of contingency as a determinant of social responsiveness. In E. Thoman (Ed.), *The origins of social responsiveness* (pp. 33–64). Erlbaum.
Weller, L., Schwarz, K. A., Kunde, W., & Pfister, R. (2018). My mistake? Enhanced error processing for commanded compared to passively observed actions. *Psychophysiology, 55*(6), Article e13057. https://doi.org/10.1111/psyp.13057
Wessel, J. R., Danielmeier, C., Morton, J. B., & Ullsperger, M. (2012). Surprise and error: Common neuronal architecture for the processing of errors and novelty. *The Journal of Neuroscience, 32*(22), 7528–7537. https://doi.org/10.1523/JNEUROSCI.6352-11.2012
Yasuda, A., Sato, A., Miyawaki, K., Kumano, H., & Kuboki, T. (2004). Error-related negativity reflects detection of negative reward prediction error. *NeuroReport, 15*(16), 2561–2565. https://doi.org/10.1097/00001756-200411150-00027