Acta Psychologica 212 (2021) 103216

Contents lists available at ScienceDirect

Acta Psychologica

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

# ERP markers of action planning and outcome monitoring in human – robot interaction

Nina-Alisa Hinz [a,b], Francesca Ciardo [a], Agnieszka Wykowska [a,*]

[a] Istituto Italiano di Tecnologia, Via Enrico Melen 83, 16152 Genova, Italy
[b] Department of Psychology and Educational Sciences, Ludwig-Maximilians-University, Geschwister-Scholl-Platz 1, 80539 München, Germany

ARTICLE INFO

*Keywords:*
Joint action
Human-robot interaction
Readiness potential (RP)
P1/N1 complex
Feedback-related negativity (FRN)

ABSTRACT

The present study aimed to examine event-related potentials (ERPs) of action planning and outcome monitoring in human-robot interaction. To this end, participants were instructed to perform costly actions (i.e. losing points) to stop a balloon from inflating and to prevent its explosion. They performed the task alone (individual condition) or with a robot (joint condition). Similar to findings from human-human interactions, results showed that action planning was affected by the presence of another agent, robot in this case. Specifically, the early readiness potential (eRP) amplitude was larger in the joint, than in the individual, condition. The presence of the robot affected also outcome perception and monitoring. Our results showed that the P1/N1 complex was suppressed in the joint, compared to the individual condition when the worst outcome was expected, suggesting that the presence of the robot affects attention allocation to negative outcomes of one's own actions. Similarly, results also showed that larger losses elicited smaller feedback-related negativity (FRN) in the joint than in the individual condition. Taken together, our results indicate that the social presence of a robot may influence the way we plan our actions and also the way we monitor their consequences. Implications of the study for the human-robot interaction field are discussed.

## 1. Introduction

Robots are beginning to be present in many aspects of our life – as virtual assistants telling us the weather forecast, or, as humanoid robots, greeting us at the airports. The "One Hundred Year Study on Artificial Intelligence (AI100)"-Report (Stone et al., 2016) predicts a further increase in robotic presence in our life, in health care as well as home environment. This makes it necessary to examine the effect of robotic presence on human behavior. In human social contexts, it has been observed that the behavior of one individual changes in the presence of other people, compared to his or her behavior in an individual situation. For example, people tend to make riskier choices (Zajonc et al., 1970) or intervene less in emergencies (Chekroun & Brauer, 2002) in the presence of others. Thus, given the role that robots might take in our society in the future, it appears crucial to investigate if the presence of robots affects individual behavior similarly to when we interact with another human.

In social contexts, we continuously need to work together with others on joint tasks and involve in collaborative actions during daily activities. However, collaboration requires planned coordination, which, according to the task and the context, relies on a combination of different cognitive mechanisms that allows the representation of one's own and others' actions and goals simultaneously (see Vesper, Butterfill, Knoblich, & Sebanz, 2010; Vesper et al., 2016 for a review). Among these mechanisms, action planning and outcome monitoring are crucial to ensure a smooth and effective collaboration in social interaction (for a review see Knoblich et al., 2011). For instance, imagine you are the hitter in a volleyball team. As a hitter, your task is to receive the ball from the setter and hit it in a way that it will not be stopped by the opponents. So you have to decide how and when to hit the ball, which requires to be able to make good decisions in a short time and to process a large amount of information at once, such as the trajectory and the speed of the ball, the position and the readiness to act of the opponent middle blocker, etc. After hitting the ball, you will continue to monitor it, to ensure that you reach the outcome you intended, i.e. score a point for your team.

Evidence showed that action planning and outcome monitoring of one's own actions is different between individual and social contexts. Sebanz et al. (2006) showed that when pairs of participants perform a "Go"/"No-go"-task in a social context, they represent not only their own

* Corresponding author.
*E-mail address:* agnieszka.wykowska@iit.it (A. Wykowska).

https://doi.org/10.1016/j.actpsy.2020.103216
Received 31 January 2020; Received in revised form 12 November 2020; Accepted 16 November 2020
Available online 4 December 2020

actions but also those of the co-actor. In the experiment of Sebanz et al. (2006), participants had to respond only to one stimulus feature (i.e. the color, “Go” trials), either individually or together. Results showed that for “No-go” trials P300 amplitude was larger in the joint than in the individual condition. Importantly, in the no-go joint condition, it was the co-actor who was producing the response. The results indicated that a representation of the action to be performed was activated following “No-go” stimulus presentation, and was then suppressed to avoid acting when it was the other’s turn. The finding that participants create a representation of co-actors’ action has been replicated in many studies (e.g., Atmaca et al., 2011; Ciardo et al., 2016; Ciardo & Wykowska, 2018; Sebanz et al., 2003), even in situations in which it is not beneficial for task performance (de Bruijn et al., 2008). Evidence shows that social context also affects visual attention. Baess and Prinz (2015) showed that the N1, an event-related potential (ERP) component indicative of early perceptual processing, was less pronounced in the joint compared to the individual action condition, although the physical features of the stimuli were similar across conditions. This implies that the social context can affect also the early stages of perceptual processing.

## 1.1. The present study

The present study aimed to examine ERP components of two mechanisms underlying joint action: action planning and outcome monitoring. Our main aim was to determine whether it is possible to identify ERP markers that characterize human-human joint action also when interacting with a robotic agent. To this end, we employed a modified version of the Balloon Risk-Taking task (BART, Beyer et al., 2018; Lejuez et al., 2002). Participants were instructed to stop a balloon from inflating, to prevent its explosion. In a within-subjects design, we manipulated the context in which participants performed the task. In one condition, participants performed the task alone (individual condition), whereas in the other case (joint condition) they performed the task with the Cozmo robot (Anki Inc., San Francisco). The only difference between the individual and the joint condition was that in the joint condition also the robot was in charge of stopping the balloon. Every possible outcome (the balloon burst, or did not burst) was associated with feedback, i.e. losing a number of points. When the robot acted instead of the participant, no points were lost.

### 1.1.1. Action planning

One of the main ERP components related to action planning is the movement readiness potential, a negative deflection over central areas starting about 2 s prior to action onset (Shibasaki & Hallett, 2006). Historically the readiness potential has been differentiated into two parts: the early part (which will be denoted here as eRP), and the later part, that is more lateralized to the side contralateral to the movement and starts about 500 ms before the movement, the lateralized readiness potential (LRP), which is related to movement preparation. While the eRP has been hypothesized to originate from the supplementary motor area (SMA) and the pre-SMA, the LRP originates in primary motor areas contralateral to an acting hand/arm (Schmidt et al., 2016). Together, the eRP and the LRP are considered two indexes of different stages of action planning (de Jong et al., 1988; Gratton et al., 1988; Leuthold & Jentzsch, 2002). While the LRP represents pure motor preparation, the eRP has been shown to be influenced by several motivational and cognitive factors, such as the complexity of the action selection process (Shibasaki & Hallett, 2006). More importantly for the aim of the present study, modulations of the readiness potential (RP) have been reported in joint action studies (Kourtis et al., 2010). For instance, Kourtis et al. (2014) investigated movement-related ERPs when participants were asked to lift their arm as if they were to clink a glass. The arm lifting occurred either in an individual or a social context. The results showed that the RP was significantly enhanced when participants were asked to coordinate with another person compared to when they were initiating the same action individually. On the other hand, social context seems to not modulate the LRP (Kourtis et al., 2014), especially if the movement to be performed alone or in the presence of the co-agent is the same in terms of complexity of kinematics and environmental constraints. To investigate action planning during interaction with a robot, we examined the modulation of the eRP as a function of social context. We expected that the eRP would be more enhanced when planning to act in the context of a joint action, as compared to planning the same action individually. To exclude that differences across conditions in the eRP could be driven by a different involvement of the primary motor areas in action planning, we also examined the LRP. Given that the action required to stop the balloon did not differ across conditions, we expected that the LRP amplitude would be comparable between the individual and the joint condition.

### 1.1.2. Outcome monitoring

Our second target of interest was outcome monitoring in interaction with a robot. When we monitor the success of our own performance, we involve a variety of cognitive mechanisms, such as perceptual and attentional processing of the outcome, and outcome monitoring per se, related to reward-based learning. In the present study, we examined both visual attention related to the outcome, and outcome monitoring. To investigate the effect of social context on ERPs markers of visual attention, we analyzed the visual P1 and N1 ERP components (Mangun et al., 1993). These are the first positive peak (P1) around 100 ms and the first negative peak (N1) around 150 ms after the presentation of a visual stimulus. They have been postulated to originate in the extrastriate occipital cortex (Luck, 2014). In a recent fMRI study, Beyer et al. (2018) found modulations of the visual cortex activity after outcome presentation in a BART task. The authors reported higher activation in the occipital cortex for the individual compared to the social condition. The authors concluded that participants paid more attention to the outcome in the individual condition and consequently monitored the outcomes of their actions less in the social context. Outcome monitoring has been mostly been correlated with the Feedback-Related Negativity (FRN), which is a negative deflection around 250 ms after the feedback presentation. It has been postulated to originate in the anterior cingulate cortex. The FRN has been hypothesized to represent processes like learning and motivation (San Martín, 2012). Therefore, it can be considered a proxy for higher-level outcome monitoring. The FRN seems to be affected by both social context and outcome valence (Beyer et al., 2017; Czeszumski et al., 2019; Li et al., 2010). For instance, Czeszumski et al. (2019) found that FRN is more negative in a competitive situation compared to a cooperative situation, as well as for negative, compared to positive, outcomes (see also Beyer et al., 2017). Evidence showed that FRN is sensitive to both expectations about the outcome and its valence, with larger FRN elicited for outcomes worse than expected (Holroyd & Coles, 2002). Moreover, outcome valence seems to affect FRN as a function of the social situation. Itagaki and Katayama (2008), using a gambling task, investigated the relationship between one’s own and others’ outcomes under cooperative or competitive instructions. Results showed that the FRN for the co-agent’s feedback was elicited by negative outcomes in a cooperative situation and by positive outcomes when in competition.

In the present study, we hypothesized that perceptual/attentional processing of outcome would be reduced in the joint condition compared to the individual condition, thus the amplitude of both P1 and N1 were expected to be smaller in the joint, compared to the individual, context, in line with previous studies (Baess & Prinz, 2015; Beyer et al., 2018). We also predicted that outcome monitoring should be affected by the social context, with the FRN amplitude being reduced for self-generated outcome in the joint, compared to the individual condition. Moreover, we expected that outcome valence would modulate the FRN amplitude, with larger FRN amplitude associated with a larger amount of lost points, hypotheses in line with previous findings (Beyer et al., 2017; Czeszumski et al., 2019).

## 2. Methods

### 2.1. Subjects

Thirty-two healthy adults took part in the study. They all provided written informed consent before participation and were debriefed after the experiment. Their participation was financially compensated. The experiment was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee (Comitato Etico Regione Liguria). The data of five participants were excluded due to technical failure of the robot (1 participant), technical failure of the EEG recording system (1 participant) or because of EEG data that was too noisy (3 participants). Therefore, data of twenty-seven participants were further analyzed. The remaining sample had an age range of 19 to 44 years (M = 23.85 years, SD = 4.81 years) and consisted of fifteen females and twelve males, one was left-handed.

### 2.2. The Cozmo robot

The interaction partner was the Cozmo robot, a commercial robot (Anki Inc., San Francisco). It is a small vehicle-shaped robot with four wheels and motors, a vertically moveable head with a facial display and camera, and a vertically moveable lift. Through the camera, it can recognize faces and also objects. It is controlled by an iOS- and Android-compatible application that can be used to program the robot also externally through a python 3.6-based software-development kit (SDK). The programming is limited to a pre-defined range of functions. In the present study the function "play_anim_trigger" with the attributes "GoToSleepSleeping", "GoToSleepGetOut", "OnSpeedtapTap" was used, as well as "go_to_object", "set_lift_height" and "set_lift_angle". Commands were sent through OpenSesame Version 3.1.9 (Mathôt et al., 2012) to the Cozmo application. Therefore, a mobile Android device with the Cozmo application was connected to the laptop through the Android Debug Bridge (adb).

### 2.3. Questionnaires

To assess the participant's attitude toward robots, before starting the experiment we administered three different questionnaires:

- The *Frankenstein Syndrome Questionnaire* (Nomura et al., 2012; Syrdal et al., 2013), measuring the "Frankenstein Syndrome" in the context of humanoid robots. The Frankenstein Syndrome is the fear of creating human-like entities that eventually might turn against their creator (Rollin, 1995). The questionnaire consists of thirty items, for which participants rate their accordance on a 7-point Likert scale. The ratings are summarized to four subscales: "General Anxiety toward Humanoid Robots", "Apprehension toward Social Risks of Humanoid Robots", "Trustworthiness for Developers of Humanoid Robots" and "Expectations for Humanoid Robots in Daily Life". While the subscales only demonstrated medium to good internal consistency, the questionnaire in total is characterized by good reliability (Syrdal et al., 2013).
- The *Negative Attitude Toward Robots Scale* (NARS; Syrdal et al., 2009) measuring the participant's attitude toward different aspects of robot use by using three subscales: "Negative Attitudes toward Situations and Interactions with Robots", "Negative Attitudes toward Social Influence of Robots" and "Negative Attitudes toward Emotions in Interaction with Robots". The questionnaire consists of fourteen items, for which participants rate their agreement on a 5-point Likert-scale. High internal consistency and validity have been demonstrated for this questionnaire (Syrdal et al., 2009).
- The *Robotic Social Attributes Scale* (Carpinella et al., 2017) measuring the social perception of robots on the scales "Warmth", "Competence" and "Discomfort". The questionnaire consists of eighteen adjectives. For each, participants need to rate on a 9-point Likert scale how much they associate it to robots in general. All the subscales have been validated psychometrically (Carpinella et al., 2017).

All questionnaires were presented on a computer screen, using OpenSesame Version 3.1.9 (Mathôt et al., 2012) and responses were made with a standard computer mouse.

### 2.4. Experimental setup and paradigm

Participants were seated at a table facing Cozmo. Between the participant and the Cozmo robot, a computer screen was placed horizontally on the table, so that both could see the screen (Fig. 1). The response devices were in-house-built keys that were mounted on top of the cubes with which Cozmo is standardly equipped. With this setup, Cozmo was able to detect the location of the cube and reach it autonomously. The keys mounted on the cubes guaranteed temporal precision of logging the tapping event.

Each trial began with the visual instruction (1000 ms) of whether the participants would be playing alone (individual condition) or together with Cozmo (joint condition) (see Fig. 2). Both conditions were presented in a block-wise fashion, in a random sequence, with nine blocks for each condition (each block including 10 trials). Subsequently to the instruction regarding the condition, a text "the trial is starting" was presented on the screen for 1500 ms. Afterwards, a fixation display was presented for a random duration between 800 and 1000 ms (the duration was chosen from this time range at the beginning of each trial and kept for all fixation displays in the respective trial). Then participants were presented with a display depicting the balloon at the starting size in the middle of the display and a pin on top of the display for 500 ms. Finally, the balloon started inflating with a speed that was variable within the trial and between trials. The task was to press the key to stop the balloon from inflating, preventing it from touching the pin and exploding. In the individual condition, only the participant could stop the balloon, whereas in the joint condition also Cozmo could stop the balloon. Cozmo was programmed to act in 60% of trials in each joint block when the balloon reached 90% of its maximal size. When the balloon was stopped, it was presented in its final size for 1000 ms. In case it exploded, a cartoon "pop" was displayed (cf. Fig. 2). In every trial, the agent (the participant or Cozmo) who stopped the balloon lost points from his/her/its initial amount of 2500 points. The larger the balloon was, the fewer points were lost. When the balloon exploded, the maximal amount of points was lost (80–100) by both players. However, the relationship between the balloon sizes and the points lost (outcome) was not linear (see Table 1). We defined four ranges of balloon size that corresponded to four ranges of outcomes. Within each of these ranges, a random value was picked, whenever the balloon was stopped. It should

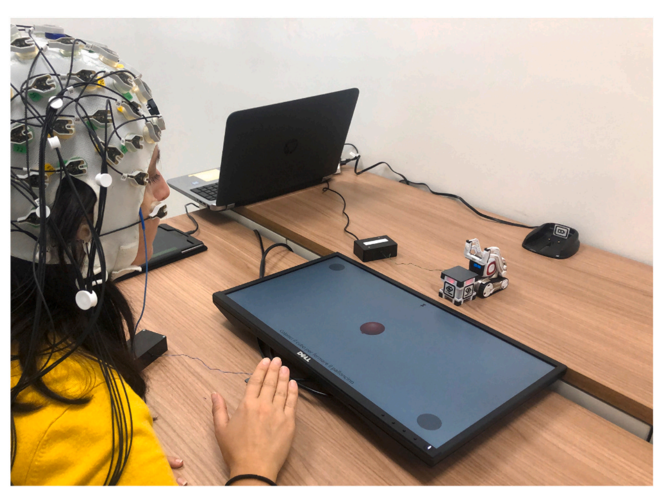

**Fig. 1.** Experimental setup.

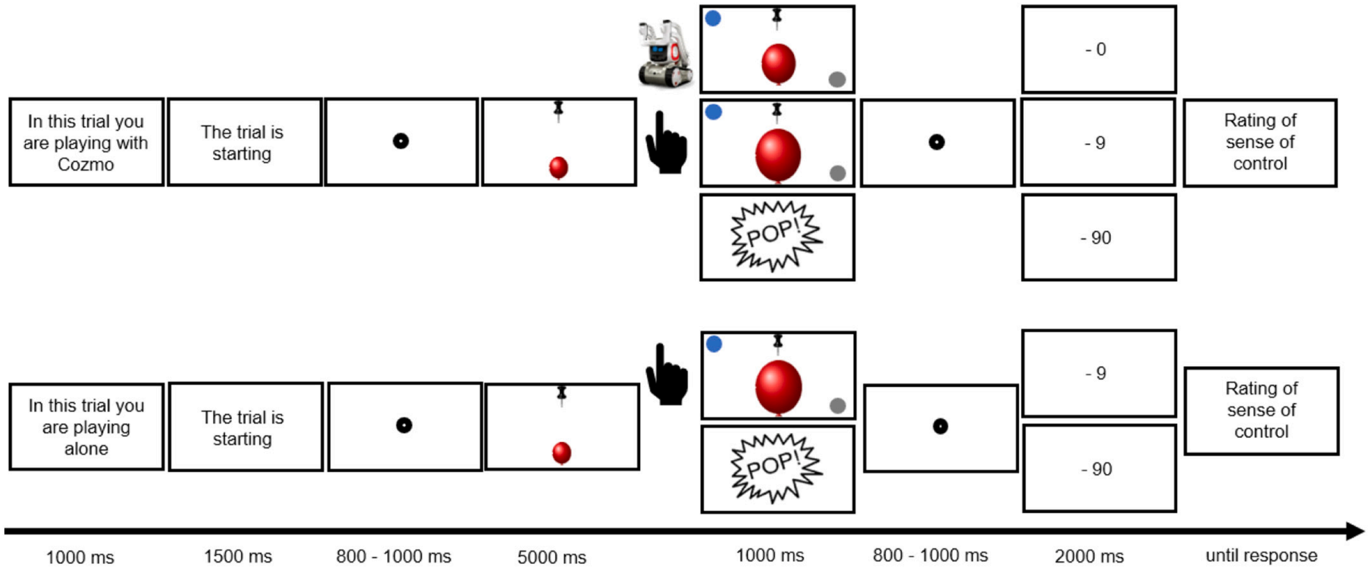


**Fig. 2.** Trial procedure in the joint condition (top panel) and in the individual condition (bottom panel). Participants and Cozmo (in the joint condition only) had to stop the balloon from inflating before it exploded. In every trial, the one who stopped the balloon lost points. The larger the balloon was at the moment of the stopping keypress, the fewer points were lost. At the end of the trial, participants were asked to indicate how much control they felt over the outcome. However, the analysis of this rating is outside the scope of this paper.

**Table 1**
The assignment of outcome ranges to balloon size ranges. The balloon size at reaction determined how many points were lost. The actual outcome was randomly drawn from the respective range of points.

| Balloon size (percentage of maximal size) | Outcome (lost points) |
|---|---|
| ≤ 17% | 46–60 |
| 17–33% | 31–45 |
| 33–49% | 16–29 |
| ≥ 50% | 1–15 |
| 100% (explosion) | 80–100 |

be noted that given the structure of the payoff, balloon size at the moment of stopping it, and the number of lost points (outcome) could be considered two different aspects of performance outcome. In our task, balloon size can be considered a proxy of the expected outcome, i.e. if the balloon is stopped at large sizes, smaller losses are expected. The number of lost points is the actual outcome, which can be considered more positive the fewer points are lost, and vice versa.

### 2.5. EEG acquisition and analysis

EEG data were recorded from 64 active Ag/AgCl electrodes, mounted on an ActiCap (Brain Products GmbH, Munich). Eye activity was recorded by one of the active electrodes placed underneath the right eye and two additional passive electrodes placed on the outer canthi of both eyes. One additional passive electrode was placed on the right ear lobe as a reference for the other two. The data were online referenced to the FCz electrode, amplified with BrainAmp amplifiers (Brain Products GmbH, Munich) and recorded with a 1000 Hz digitization rate by the BrainVision Recorder (Brain Products GmbH, Munich). No online filters were applied. Impedances of the active electrodes were controlled to be lower than 10 kΩ.

The data were pre-processed with EEGLAB version 14.1.2. (a MATLAB-based open-source toolbox; Delorme & Makeig, 2004). Raw data were re-sampled to 250 Hz and re-referenced to the averaged mastoids (TP9, TP10). Data on the FCz electrode were interpolated. A notch filter (FIR filter, 50 Hz) was applied to reduce line noise. The subsequent pre-processing steps depended on the time epochs of interest: for analysis of action planning-related data, a 0.05 Hz high-pass filter (e.g. Reznik et al., 2018) was applied, while for analysis of outcome monitoring-related data a 0.1 Hz high-pass filter (e.g. Beyer et al., 2017; Czeszumski et al., 2019) was used.[1] The data then were segmented into 3200 ms epochs for action planning-related data (from 2700 ms before the response to 500 ms after the response, the baseline-correction was done with the first 200 ms of the epoch) and 3000 ms epochs for outcome monitoring-related data (from 500 ms before the outcome presentation to 2500 ms after it; baseline: 100 ms before the presentation; as in Beyer et al., 2017). Independent component analyses were conducted on each data set separately to remove horizontal eye-movement artifacts and blinks. After that, a low-pass filter was applied (cutoff frequency 40 Hz for action planning-related data, cutoff frequency 20 Hz for outcome monitoring-related data (Reznik et al., 2018) and epochs were semi-automatically rejected based on an 80-μV threshold. This resulted in 7.33% removed trials for the action planning-related data and 11.13% removed trials for the outcome monitoring-related data.

ERPs were analyzed with FieldTrip version 2019/02/09 (a MATLAB-based open-source toolbox; Oostenveld et al., 2011). For action planning-related ERPs, we analyzed the eRP and the LRP. The eRP was defined as the mean amplitude over the electrodes Fz, FCz, Cz in the time period of 2500 ms to 500 ms preceding the response onset. Since responses in this experiment were only given with the right hand, the LRP was calculated by subtracting activity in electrode C4 from electrode C3 (C3–C4). The amplitude of the LRP was then defined as the mean amplitude from this difference in the time range of 500 ms before the response onset. For outcome monitoring, early visual attention-related potentials (P1 and N1) and the later FRN were analyzed. The early visual attention-related ERPs were analyzed as an average of the electrodes Oz, O1, O2, PO7, and PO8. The P1 amplitude was defined as the

[1] A high-pass filter of 0.1 Hz is highly recommended for the analysis of EEG-data (Luck, 2014). However, when looking at components as the LRP, which are also called slow-cortical potentials (Schmidt et al., 2016) filters of 0.1 Hz cut off many of the slow-waves that actually contribute to the LRP and therefore can distort the results (Luck, 2014). To make sure that both components which are very different are analyzed correctly and data are not distorted, we applied different filters. The filters we used were based on literature, as referenced above.

most positive value between 90 and 140 ms after outcome-presentation (Luck, 2014), while the N1 amplitude was the most negative value between 130 and 190 ms post-onset (Luck, 2014). The FRN was defined as the mean amplitude over electrodes Cz, C1, C2, FCz, FC1, FC2, Fz, F1, and F2 in the time range between 200 and 300 ms after the outcome presentation. Although the FRN is usually extracted from difference waves between positive and negative outcomes, previous studies also looked at the FRN in trials with outcomes of only one valence (i.e. either positive or negative) (e.g. Beyer et al., 2017; Czeszumski et al., 2019). Also in this study, the outcomes were only of negative valence and thus the FRN is not analyzed as a difference between positive and negative outcomes. For the additional exploratory peak-to-peak analysis with the FRN, we defined the positive peak preceding the FRN as the most positive amplitude over electrodes Cz, C1, C2, FCz, FC1, FC2, Fz, F1, and F2 in the time range between 100 and 200 ms after the outcome presentation. We then subtracted this amplitude from the FRN amplitude to calculate the peak-to-peak difference.

### 2.6. Statistical analysis

Firstly, we assessed how the behavior of participants was altered by the presence of the robot, thus we analyzed how often participants or Cozmo successfully stopped the balloon. The proportions of reactions of each agent or explosions were compared with paired-sample *t*-tests. Subsequently, only trials in which participants stopped the balloon were analyzed, given that in these trials, the only difference across the conditions of interest (individual vs. joint) was the social context.[2]

Linear mixed effect models were used to analyze the behavioral and electrophysiological data. Outcomes and balloon sizes were z-transformed within each participant. The dependent variables were amplitudes of the eRP, LRP, P1, N1 and FRN components and the peak-to-peak difference of the FRN and the preceding positive peak. Fixed effects were the condition, z-transformed balloon sizes and outcomes (only in the P1, N1, FRN, and peak-to-peak difference model), and their interactions. For each participant we modelled a random intercept. Additional covariates as sex (fixed effect) or the scores from the questionnaires (total scores of the FSQ, NARS, and RoSAS questionnaires as random intercepts) were included on a step-by-step basis and only considered if a likelihood ratio test of the model with effect in question against the model without the respective effect became significant. *p*-Values on the estimates were obtained by an ANOVA on the model. All analyses were conducted using R Version 3.5.1 (R Core Team, 2018) and the lme4 package Version 1.1.81.1 (Bates et al., 2015). The ggplot2 package Version 3.0.0. (Wickham, 2016) was used to create plots. All model estimates, as well as confidence intervals, are reported in the Supplementary material (SM, Section 10.2).

## 3. Results

The mean total score of the NARS was 32.57 ($SD = 5.68$). The mean total score of the FSQ was 108.81 ($SD = 12.43$). The mean total score of the RoSAS was 77.45 ($SD = 13.94$). The means and standard deviation of the subscales of all questionnaires are reported in Table 2.

Participants successfully stopped the balloon in 87.42% ($SD = 7.9\%$) of all trials in the individual condition (see Fig. 3). Therefore, the balloon exploded in 12.58% ($SD = 7.9\%$) of trials in the individual condition. In the joint condition the number of explosions decreased significantly ($t_{26} = -2.71$, $p = .01$, $d = 0.52$) to 9.26% ($SD = 4.1\%$). Participants let Cozmo act in 39.48% ($SD = 6.9\%$) of trials, while they stopped the balloon themselves in 51.60% ($SD = 8.9\%$) of trials. This percentage differed significantly from the percentage of reactions in the individual condition ($t_{26} = -21.27$, $p < .001$, $d = 4.09$). All additional behavioral analyses are reported in the Supplementary Material (SM, Section 10.1).

**Table 2**
Mean scores and standard deviations for the subscales of the NARS, FSQ, and RoSAS.

| Questionnaire | Scale | M | SD |
|---|---|---|---|
| NARS | Negative attitudes toward situations and interactions with robots | 10.45 | 3.43 |
| | Negative attitudes toward social influence of robots | 11.97 | 3.94 |
| | Negative attitudes toward emotions in interaction with robots | 10.15 | 2.72 |
| FSQ | General anxiety toward humanoid robots | 36.40 | 11.58 |
| | Apprehension toward social risks of humanoid robots | 24.03 | 3.74 |
| | Trustworthiness for developers of humanoid robots | 21.89 | 3.45 |
| | Expectations for humanoid robots in daily life | 26.48 | 5.60 |
| RoSAS | Warmth | 17.91 | 8.34 |
| | Competence | 44.77 | 6.41 |
| | Discomfort | 14.76 | 6.78 |

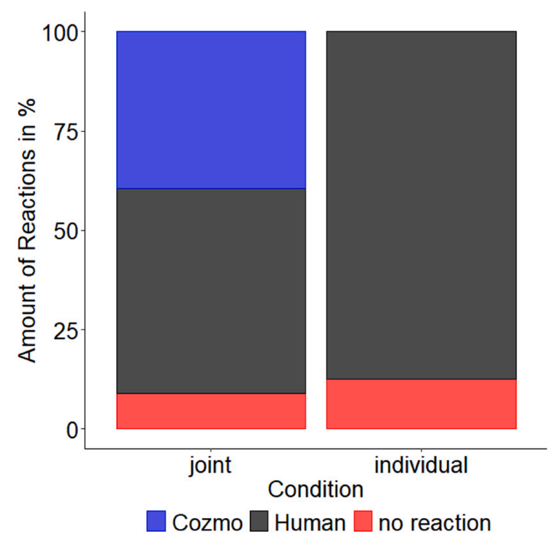


**Fig. 3.** Proportions of reactions in each condition. In the case of no reaction, the balloon exploded.

### 3.1. Action planning-related ERPs

#### 3.1.1. Early readiness potential

The eRP model was significantly improved by adding the FSQ score ($X^2 = 5.20$, $p = .02$; for an exploratory analysis on this effect see the SM), therefore estimates from this model are reported (see SM Table 6). No other covariate improved the model (all *ps* $> .49$). The social context significantly predicted the eRP amplitude ($b = -0.69$, $F_{26.89,1} = 5.22$, $p = .03$), with more negative amplitude in the joint condition ($M = -1.65$ μV, $SD = 2.56$ μV; see Fig. 4) than in the individual condition ($M = -1.01$ μV, $SD = 2.00$ μV), but the balloon size did not ($b = 3.42$, $F_{34.89,1} = 0.02$, $p = .88$). The interaction between condition and balloon size marginally affected the eRP amplitude ($b = -7.17$, $F_{14.21,1} = 4.08$, $p = .06$). Consequently, the eRP amplitude was slightly more positive for larger balloon sizes in the individual condition, while it was less positive for larger balloon sizes in the joint condition (see Fig. 5).

[2] In contrast, the no-action trials might have differed with respect to more than one aspect. In the individual condition, when participants did not prevent the balloon from bursting, they decided to not act (or they were too slow). However, in the joint condition, they might have either decided to not act, or they might have been too slow, or they might have decided to wait for Cozmo. Hence, the decision processes that led to “non-action” varied across conditions on multiple aspects.

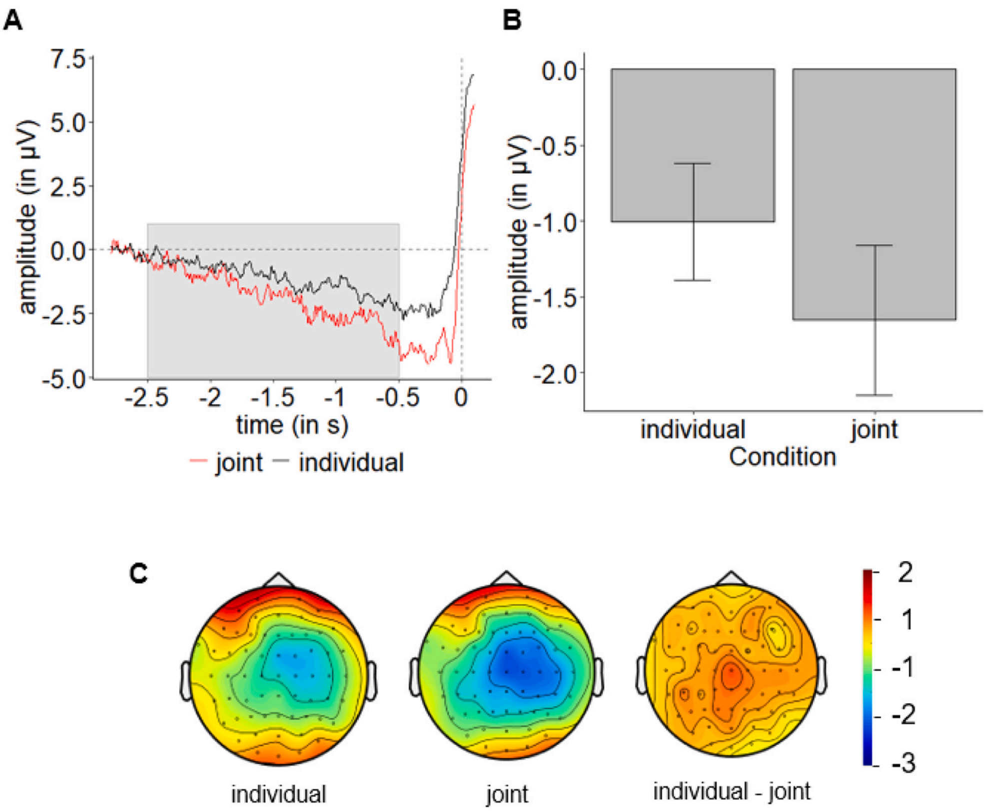


**Fig. 4.** Early readiness potential over electrodes Fz, FCz, Cz in the time window between −2.5 s and −0.5 s. Grand average waveforms in both conditions (A). Mean amplitude and standard error in both conditions (B). Scalp topographies in both conditions and the difference between them (C). The scalp topographies for the 2500 ms preceding the response.

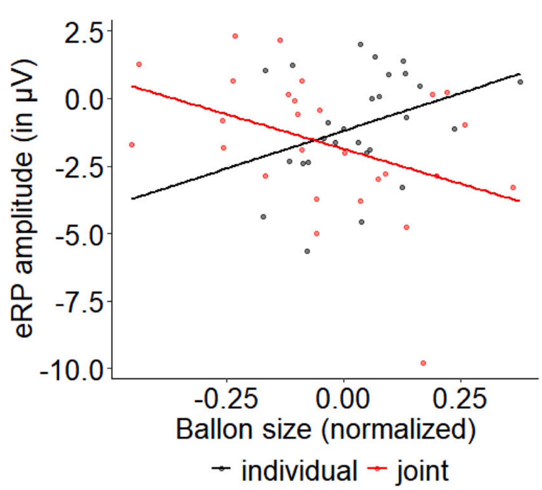


**Fig. 5.** The interaction effect between balloon size (normalized values) and condition, predicting the early readiness potential amplitude.

### 3.1.2. *Lateralized readiness potential*

The LRP model was significantly improved by adding the FSQ score ($X^2 = 6.98$, $p < .001$; for an exploratory analysis on this effect see the SM), therefore estimates from this model are reported (see SM Table 7). No other covariate further improved the model (all *ps* > .23). The social context did not have an influence on the LRP amplitude ($b = -0.49$, $F_{27.11,1} = 2.83$, $p = .10$), amplitudes were similar in the individual condition (M = 1.02 μV, SD = 3.48 μV; see Fig. 5) and in the joint condition (M = 0.62 μV, SD = 3.57 μV). Neither the balloon size ($b = -0.57$, $F_{35.51,1} = 1.02$, $p = .32$) nor the interaction between condition and balloon size ($b = -1.04$, $F_{10.19,1} = 0.05$, $p = .82$) were significant predictors of the LRP amplitude. Please note that also the LRP itself was not significantly different from zero, suggesting that the present experiment did not elicit an LRP at all (Individual: $t_{26} = 1.52$, $p = .14$; Joint: $t_{26} = 0.90$, $p = .38$).

## 3.2. *Outcome monitoring-related ERPs*

### 3.2.1. *P1*

The P1 amplitude (see Fig. 6) was significantly affected by the social context ($b = -1.33$, $F_{26.15,1} = 9.43$, $p = .005$), with more positive amplitude in the individual ($M = 2.98$ μV, $SD = 3.34$ μV) than in the joint condition ($M = 1.91$ μV, $SD = 1.89$ μV). In addition, the balloon size had a significant effect on the P1 amplitude ($b = -12.37$, $F_{41.69,1} = 6.03$, $p = .02$), while the effect of the outcome was only marginal ($b = -6.86$, $F_{41.64,1} = 3.31$, $p = .08$). The interaction between condition and balloon size reached the level of significance ($b = 15.87$, $F_{28.63,1} = 6.95$, $p = .02$; all other *ps* > .14), implying that balloon sizes only had an effect on the P1 amplitude in the individual condition but not in the joint condition, as the estimated effect of size of the balloon in the joint condition was not significantly different from zero ($b = 3.5$, 95% CI = [−1.23, 8.36], see Fig. 7). Specifically, for the individual condition, smaller balloon sizes predicted more positive amplitudes. None of the covariates significantly improved the model (all $p > .21$, see SM Table 8).

*N1*. The N1 model was significantly improved by adding the FSQ score ($X^2 = 6.98$, $p < .001$; for an exploratory analysis on this effect see the SM), therefore estimates from this model are reported (see SM Table 9). No other covariates improved the model significantly (all $p < .19$). The N1 amplitude (see Fig. 8) was significantly affected by the social context ($b = -1.22$, $F_{26.48,1} = 12.77$, $p = .001$), with more negative amplitude in the joint condition ($M = -2.67$ μV, $SD = 2.09$ μV)

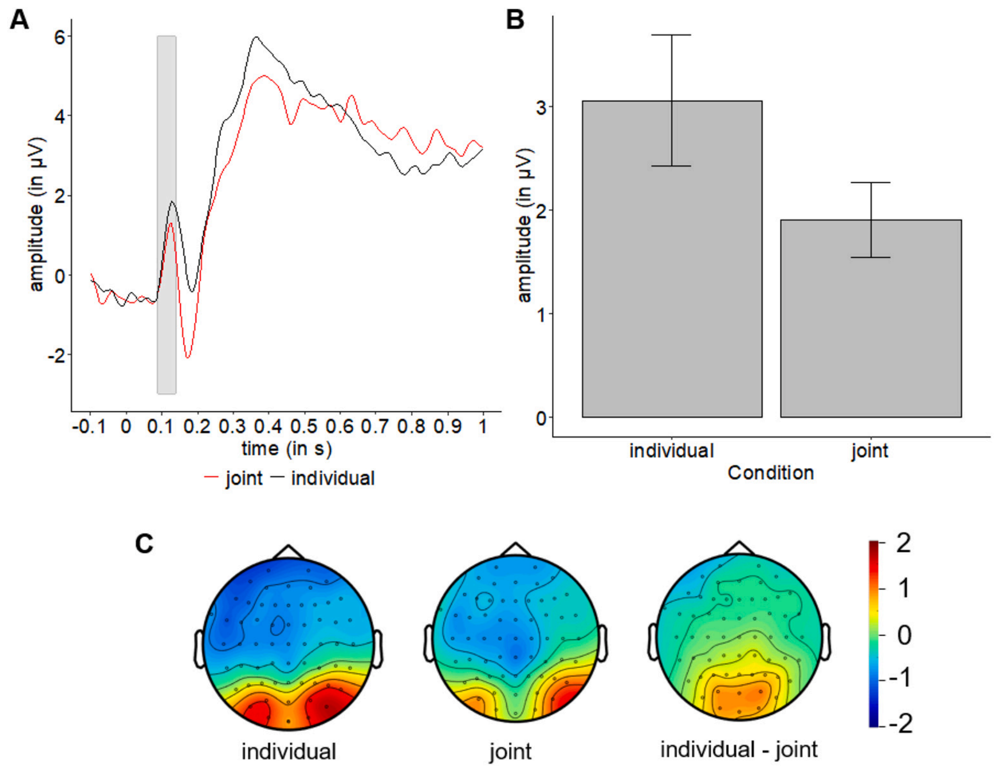


**Fig. 6.** The P1 amplitude over the electrodes Oz, O1, O2, PO7 and PO8 in the time window between 90 and 140 ms. Grand average waveforms in both conditions (A). Mean amplitude and standard error in both conditions (B). Scalp topographies in both conditions and the difference between them (C).

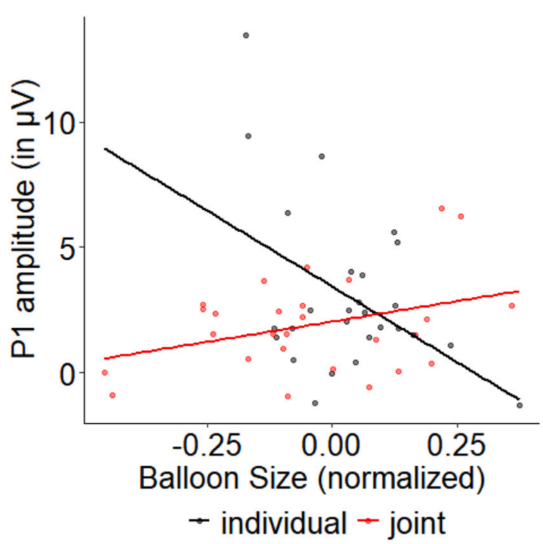


**Fig. 7.** The interaction of condition and balloon size (normalized values) on the P1 amplitude.

than in the individual condition ($M = -1.39$ µV, $SD = 2.46$ µV).

The outcome ($b = -19.68$, $F_{43.08,1} = 9.80$, $p = .003$) was a significant predictor of the N1 amplitude, while the balloon size was not ($b = -0.49$, $F_{34.86,1} = 0.47$, $p = .50$). In addition, the interaction between condition, balloon size and outcome reached the level of significance ($b = -79.58$, $F_{39.11,1} = 7.42$, $p = .01$, cf. Fig. 9). The interaction between outcome and balloon size affected N1 amplitude in the individual condition ($b = 72.36$, 95% CI = [8.63, 133.97]) but not in the joint condition ($b = -7.28$, 95% CI = [−37.45, 22.12]). No other interaction had an effect (all $p > .13$).

### 3.2.2. Feedback-related negativity

The FRN model was significantly improved by adding the ROSAS score ($X^2 = 7.64$, $p = .006$; for an exploratory analysis on this effect see the SM), therefore estimates from this model are reported (see SM Table 10). None of the other covariates significantly influenced the FRN model (all $ps > .36$). The FRN amplitude was not significantly different between joint and individual contexts (joint: M = 5.26 µV, SD = 4.26 µV, Individual: M = 5.11 µV, SD = 4.52 µV; $b = 0.19$, $F_{30.08,1} = 0.31$, $p = .58$; see Fig. 10). Outcome had a significant effect on the FRN amplitude ($b = -24.51$, $F_{35.28,1} = 5.02$, $p = .03$), with less negative amplitude for larger outcomes. Also, the interaction between social context and outcome was significant ($b = 35.84$, $F_{44.40,1} = 9.97$, $p = .003$): while the FRN amplitude was larger (i.e. more negative) for larger losses in the individual condition, in the joint condition the amplitude was smaller (i.e. more positive) for larger losses (see Fig. 11). Neither size nor any of the interactions significantly affected the FRN amplitude (all $ps > .36$).

As the positive peak preceeding the FRN seemed to show an effect based on the social context in Fig. 10, we examined – in an exploratory fashion – the peak-to-peak difference between the most positive point in the interval of 100 and 200 ms and the FRN. We ran the same models as for the FRN. None of the covariates significantly improved the model (all $ps > .32$), therefore estimates from the model without covariates are reported (see SM Table 11). The peak-to-peak difference was significantly different between conditions (Individual: $M = 8.43$ µV, $SD = 3.12$ µV; Joint: $M = 9.48$ µV, $SD = 3.63$ µV; $b = 1.02$, $F_{26.50,1} = 6.33$, $p = .02$). The effect of outcome was only marginal ($b = -10.80$, $F_{50.47,1} = 2.97$, $p = .09$). There also was a significant interaction effect between condition and balloon size ($b = 17.85$, $F_{29.38,1} = 4.79$, $p = .04$), indicating that balloon size had different effects on the amplitude depending on the social context. While in the joint condition larger balloon sizes led to a larger peak-to-peak difference, in the individual condition larger balloon sizes led to a smaller peak-to-peak difference. None of the other effects was significant (all $ps > .18$).

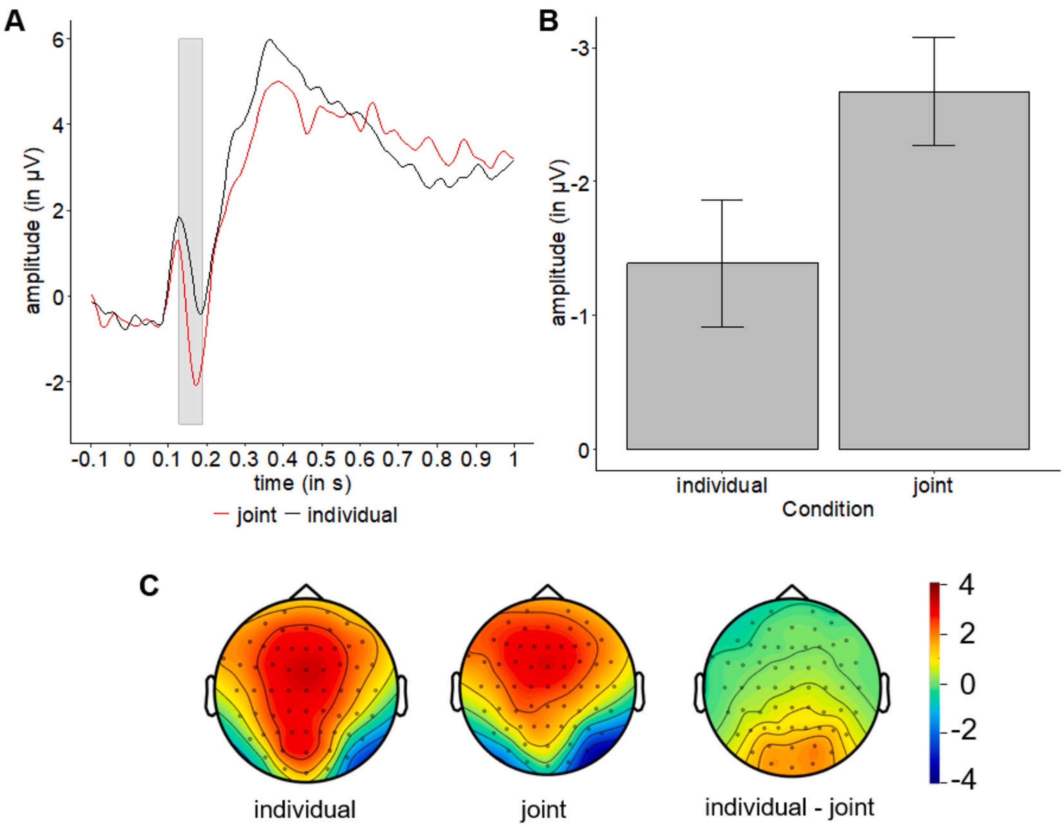


**Fig. 8.** The N1 amplitude over the electrodes Oz, O1, O2, PO7 and PO8 in the time window between 130 and 190 ms. Grand average waveforms in both conditions (A). Mean amplitude and standard error in both conditions (B). Scalp topographies in both conditions and the difference between them (C).

## 4. Discussion

The present study aimed at examining ERP markers of action planning and outcome monitoring when interacting with a robotic agent. We focused on the eRP and LRP, as ERP components related to action planning, known to be modulated by the social context in human-human interactions (Kourtis et al., 2010; Kourtis et al., 2014). Moreover, we focused on the visual attention-related P1 and N1, locked to outcome presentation. The visual attention ERPs related to the outcome would represent outcome processing. Finally, we also focused on the FRN, representing outcome monitoring.

### *4.1. Action planning*

Electrophysiological results indicate that the context in which the task was performed affected action planning at a neural level. Our results showed that the eRP was more negative (i.e. larger amplitude) for actions in the joint than in the individual condition. An increase in the RP amplitude has been related to a more demanding action selection (Shibasaki & Hallett, 2006). This seems to be a plausible explanation for the present effect, given that in the joint condition participants could act, let the balloon explode, or let the robot act. In the individual condition, however, there was only a binary decision between acting, or letting the balloon explode. Furthermore, in the individual condition participants had to act to prevent very negative outcomes and consequently, the decision was biased toward acting. In the joint condition, in contrast, to avoid negative outcomes, participants could also decide to not act and trust the robot to act instead. Thus, the action selection in the joint condition was more demanding because participants represented not only their own actions but also those of the robot. Since there was no difference between conditions in the LRP, it is evident that the eRP difference was not related to action preparation per se. Our findings are consistent with this interpretation and confirm that the modulatory effect of social context on the eRP is not driven by a differential processing at the action preparation stage, but rather by action selection processes (Kourtis et al., 2014). Finally, our results extend previous evidence by Sebanz et al. (2006) showing that interacting with a robot seems to influence action planning in a similar manner as interacting with a human partner.

### *4.2. Outcome monitoring*

#### *4.2.1. P1/N1 complex*

Regarding outcome monitoring, in line with our hypotheses, we found an effect of social context on outcome processing at the perceptual/attentional level, as indicated by modulation of both visual attention ERPs, the P1 and N1. As predicted, P1 showed larger amplitudes in the individual than in the joint condition, suggesting that social context affects early perceptual processing of the outcome, with less visual attention resources allocated to self-generated outcome in the joint, compared to the individual condition. This is in line with the findings of Beyer et al. (2018) and Baess and Prinz (2015). Interestingly, the effect of social context on P1 amplitude was modulated by the expectations about the outcome, as indicated by the significant interaction with the size at which the balloon was stopped. Expected larger losses (i.e. smaller balloon size) were associated with a more positive P1 in the individual but not in the joint condition.

Hence, it seems that the larger the expected losses, the more attentional resources were allocated to the outcome, independent of the actual outcome value. Importantly to note, the P1 was modulated by *expected* losses and not by the actual *outcome*, as there was no interaction with the outcome values. The modulation of P1 by the expected outcome was true only for the individual, not for the joint condition. This might suggest some indication of “diffusion of responsibility” (Bandura, 1991, 1999), meaning that in the social context condition, participants paid less attention to the outcome, even if they expected large losses.

Our second variable of interest in outcome processing was the N1 component. Results showed that N1 amplitude was larger (i.e. more negative) in the joint, compared to the individual condition. However, the N1 should be interpreted in relation to the three-way interaction of condition (individual vs. joint), outcome and balloon size (cf. Fig. 9) modulating the N1 amplitude. The left panel of Fig. 9 represents the individual condition, which is the baseline. The black spotted line shows that the N1 amplitude was modulated by the outcome, but more for small balloon sizes than large balloon sizes, as the effect is smaller for the condition represented by the solid black line (large balloon sizes). This suggests that in the individual condition, more attentional resources were activated (more negative N1 amplitude) when the outcome confirmed expected large losses, as compared to when the outcome did not confirm the expected large losses (in this case, N1 amplitude seems to be suppressed, compared to other conditions). On the other hand, when participants did not expect large losses (large balloon sizes, solid black line), there was less attentional modulation of outcome processing. Similarly, when participants were in the joint condition (Fig. 9, right panel), the outcome did not modulate the allocation of attention, and this was the case, independent of expectations (balloon size). Therefore, it seems that only the individual condition elicited enhancement of attention for the more negative (and expected) outcome, and suppression of attentional resources for the less negative outcome, as compared to expectations in the joint condition. This again speaks in favour of diffusion of responsibility in the joint condition, as attention is not modulated by outcome processing or expectations. Therefore, it seems that less attentional engagement toward outcome occurred in general for the joint condition, independent of the outcome value or expectations. In the individual condition, on the contrary, participants were more sensitive to the feedback they received.

Together, the results on P1 and N1 can be interpreted as a marker of monitoring success. When the expected losses were small, i.e. the balloon was stopped close to the pin (Fig. 9 solid lines), there was no difference between conditions in attentional resources allocated to the outcome. However, for actions that stopped the balloon at an early stage of its inflation (Fig. 9 spotted lines), i.e. when participants expected a larger loss, attentional resources related to outcome processing were modulated by the social context (i.e. with P1 amplitude modulated by balloon size in the individual but not in the joint condition, Fig. 7) and whether the outcome confirmed the expectations or not (N1, Fig. 9). This pattern could be interpreted as a consequence of the diffusion of responsibility (Bandura, 1991, 1999), i.e. the tendency to feel less responsible for the consequences of our actions when in a social context. Negative outcomes for self-generated actions may lead to less arousal in the social than individual context. Thus, as a consequence participants may have been less attentionally engaged in the outcome of their actions in the social, as compared to the individual condition.

Expecting worse outcomes (and confirmation of these expectations) in the individual condition probably led to more arousal and consequently increased attentiveness. The fact that this effect was smaller or even reversed in the joint condition, might point to the fact that participants monitor outcomes of their performance less in the joint condition as less relevant. Indeed, studies concerning outcome processing have shown that negative outcomes are evaluated as being less negative in group situations than the same outcome in an individual condition (Beyer et al., 2017; Li et al., 2010).

### *4.2.2. FRN*

The influence of social context on outcome monitoring was not reflected in the FRN amplitude. Such a result is not entirely in line with previous studies showing that in a similar risk-taking task the FRN is reduced in the social context compared to the individual context (Beyer et al., 2017; Czeszumski et al., 2019). However, the exploratory peak-to-peak analysis showed that this effect might be driven by context effects on the positive peak before the FRN. The positive peak could be representing the P200, a component that has been associated with attention selection and action monitoring, similar to the FRN (Potts, 2004; Potts et al., 2006). Future studies should examine these two components separately to disentangle the effects of social context on each of the components.

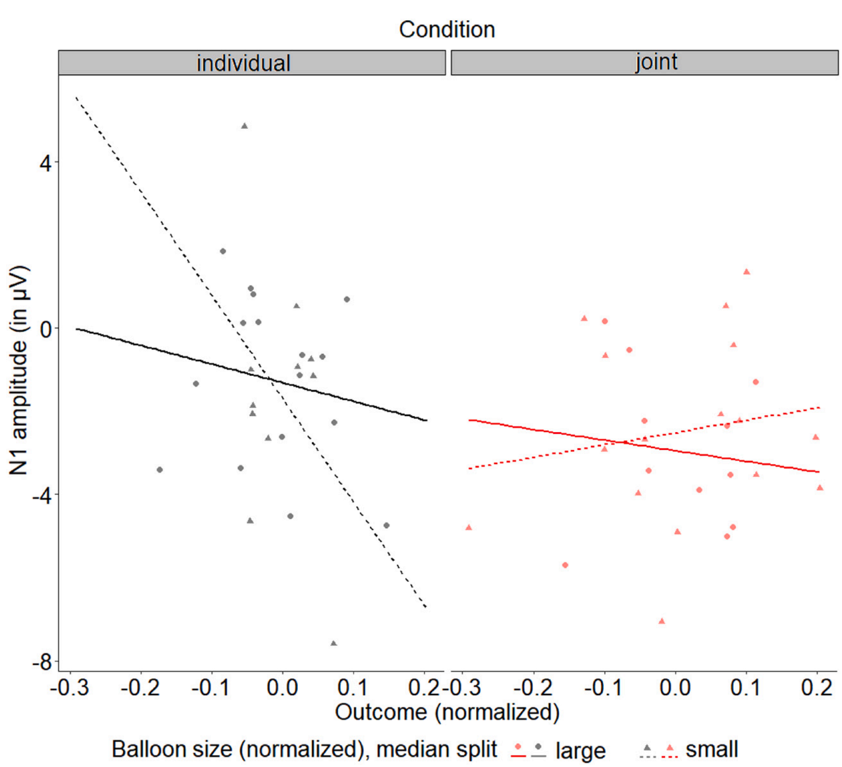


**Fig. 9.** The interaction of condition, outcome (normalized values) and balloon size (normalized values) on the N1 amplitude.

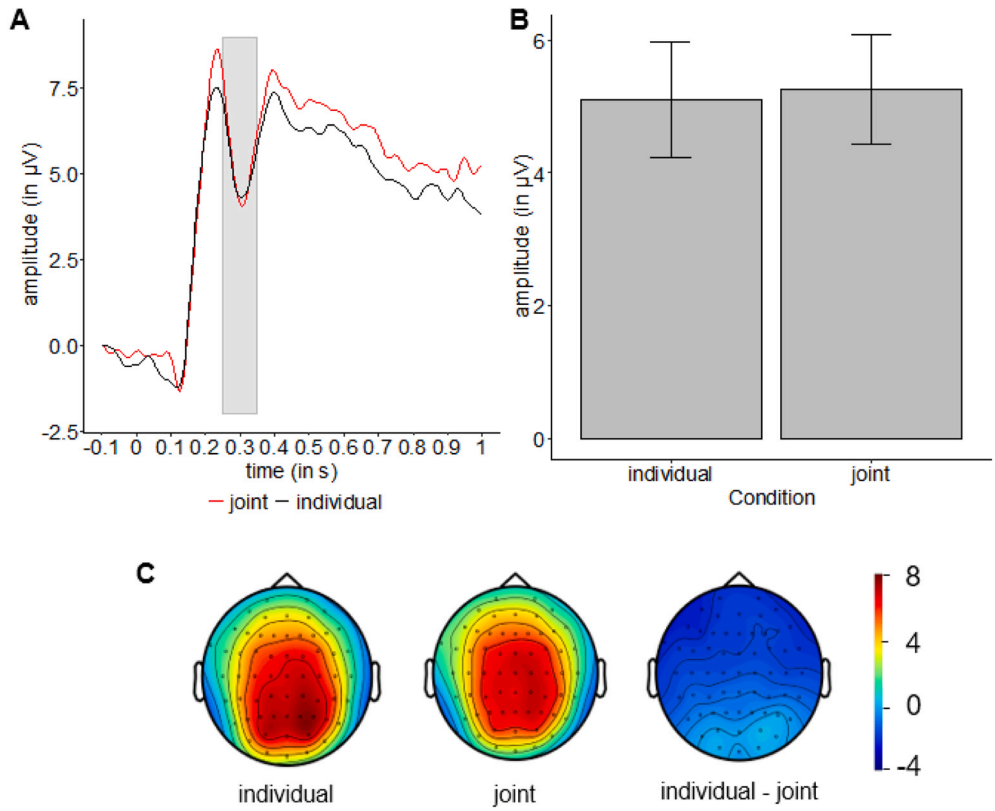


**Fig. 10.** The FRN amplitude over the electrodes Cz, C1, C2, FCz, FC1, FC2, Fz, F1 and F2 in the time window between 200 and 300 ms. Grand average waveforms in both conditions (A). Mean amplitude and standard error in both conditions (B). Scalp topographies in both conditions and the difference between them (C).

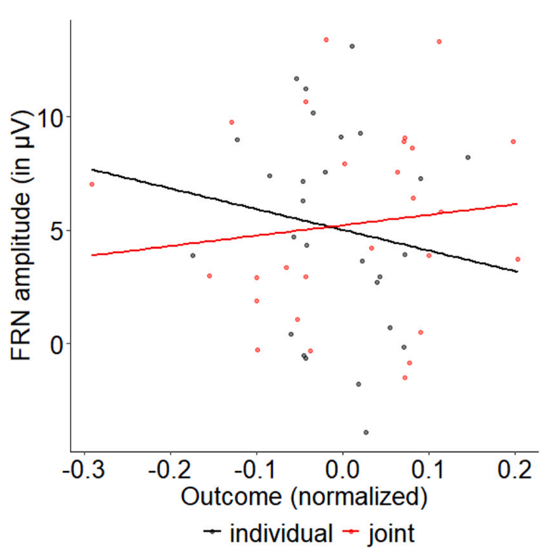


**Fig. 11.** The interaction effect of condition and outcome (normalized values) on the FRN amplitude.

Important for the current analyses and hypotheses, however, there was an interaction between outcomes and social context on the FRN amplitude, with smaller amplitudes for large losses in the joint condition, while in the individual condition it was smaller for smaller losses. Consequently, larger losses in the joint condition elicited fewer resources deployed to outcome monitoring than in the individual condition. The FRN has been hypothesized to represent reinforcement learning, the use of the presented outcome to modulate subsequent behavior (San Martín, 2012). Our results suggest that in interaction with a robot, self-generated negative outcomes (i.e. larger loss of points) might be less relevant for affecting subsequent behavior compared to when the same negative outcome occurs in the individual situation. Czeszumski et al. (2019) reported a similar pattern of results for human-human interactions. The authors showed that the FRN component is sensitive not only to positive and negative outcomes but also to the social context in which two human agents actively interact. Our results extend previous findings by showing that the interaction between outcome valence and social situation affects FRN also when the co-agent is a robot. Such a result might be a proxy of a self-serving bias (Bandura, 1999) when in interaction with the robot. Self-serving bias is the tendency to attribute to external events the responsibility of negative outcomes to maintain self-esteem (Bandura, 1999). When in social interactions, self-serving bias occurs when people believe they are interacting with an intentional agent (Beyer et al., 2017; Beyer et al., 2018; Ciardo, Beyer, De Tommaso, & Wykowska, 2020). In our study, although we did not manipulate beliefs about how the robot was controlled, it could be that participants attributed a certain level of intentional agency to it (see Marchesi et al., 2019). Thus the self-serving bias would be a consequence of mentalizing processes. However, a self-serving bias also occurs when the locus of the attribution is just the context itself (e.g., Kestemont et al., 2015), thus it is possible that framing the task as social was enough to induce the bias. Future studies should systematically address this issue examining the relationship between intentionality attribution toward robots and the occurrence of the self-serving bias.

In summary, our results from ERPs of outcome monitoring suggest that both outcome processing and monitoring were influenced by the robot presence. However, results also indicated a complex interplay between expectations and outcome valence on the modulatory effect elicited by the social presence of the robot.

### 4.3. Considerations for studying human-robot interaction

In the present study, many of the electrophysiological models were improved by accounting for individual differences in the attitude participants had toward robots, as most of the questionnaires explained a substantial amount of variance, although not indicating in which direction. It could be that the effect exerted by the social presence of Cozmo at the neural level varies as a function of individual attitude toward robots, with higher scores on the FSQ questionnaire (indicating more negative attitudes) being associated with lower differences between the individual and the social context (see Hinz et al. (2019) for preliminary behavioral results in this direction). A critical aspect when interacting with an artificial agent is the attribution of intentionality to the robot. Intentionality is a concept that is critical for the attribution of responsibility (Frith, 2014) and that is not always naturally attributed to robots (Perez-Osorio & Wykowska, 2019). Previous studies that investigated the link between joint action and intentionality attribution to robots showed that action planning processes were affected by the social presence of a robotic co-agent only when participants believed that the robot was controlled by a human (Stenzel et al., 2012). It should be noted that in the present study we used a non-anthropomorphic robot (for a possible relationship between anthropomorphism and intentionality attribution see Epley et al., 2007), and we did not manipulate beliefs regarding how the robot was controlled. Thus, any form of intentionality attributed to Cozmo emerged from a spontaneous attitude of the participants. However, we did not measure intentionality attribution directly, thus we cannot draw any conclusions about how intentionality attribution may have affected action planning and outcome monitoring in interaction with the robot. Future studies should systematically investigate if the social presence of the robot differently affects joint action as a function of several factors, such as attribution of mental states (Marchesi et al., 2019), the tendency toward anthropomorphism (Waytz et al., 2010) or toward adopting the intentional stance toward robots (Bossi et al., 2020), and skepticism toward artificial intelligence (Syrdal et al., 2009). This appears crucial concerning the results of outcome monitoring. Indeed, overall our results suggest that when interacting with a robot, both outcome processing, and monitoring are affected by the presence of the robot.

This is an important phenomenon in the context of future scenarios where robots will be present in our daily life. Indeed, the social presence of the robot may influence the ability to process our action outcome. Therefore, robots should be designed taking into account that their social presence may affect humans' decision-making, action planning at different stages, as well as learning from past action outcomes.

## Authors contribution

NAH conceived, designed, and performed the study; collected and analyzed the data, discussed and interpreted the results; wrote the manuscript. FC conceived and designed the study, discussed and interpreted the results; wrote the manuscript. AW conceived and designed the study; discussed and interpreted the results; wrote the manuscript. All authors reviewed the manuscript.

## Additional information

Behavioral (but not EEG) data of this sample was published in the proceedings of the 11th International Conference on Social Robots, Madrid, 26–28 November 2019: Hinz, N. A., Ciardo, F., & Wykowska, A. (2019, November). Individual differences in attitude toward robots predict behavior in human-robot interaction. In International Conference on Social Robotics (pp. 64–73). Springer, Cham, with a different focus and different analyses.

The data related to this study can be accessed under this link: https://osf.io/2zf86.

## Acknowledgements

This work received support by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant awarded to AW, titled "InStance: Intentional Stance for Social Attunement." G.A. No: ERC-2016-StG-715058). The content of this paper is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2020.103216.

## References

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research, 211*(3–4), 371–385. https://doi.org/10.1007/s00221-011-2709-9

Baess, P., & Prinz, W. (2015). My partner is also on my mind: Social context modulates the N1 response. *Experimental Brain Research, 233*(1), 105–113. https://doi.org/10.1007/s00221-014-4092-9

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 248–287. https://doi.org/10.1016/0749-5978(91)90022-L

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*(3), 193–209.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1). doi:10.18637/jss.v067.i01.

Beyer, F., Sidarus, N., Bonicalzi, S., & Haggard, P. (2017). Beyond self-serving bias: Diffusion of responsibility reduces sense of agency and outcome monitoring. *Social Cognitive and Affective Neuroscience, 12*(1), 138–145. https://doi.org/10.1093/scan/nsw160

Beyer, F., Sidarus, N., Fleming, S. M., & Haggard, P. (2018). Losing control in social situations: How the presence of others affects neural processes related to sense of agency. *ENeuro, 5*(1). https://doi.org/10.1523/ENEURO.0336-17.2018

Bossi, F., Willemse, C., Cavazza, J., Marchesi, S., Murino, V., Wykowska, A. (2020). The human brain reveals resting state activity patterns that are predictive of biases in attitudes towards robots. Science Robotics, 5:46, eabb6652.

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). *The robotic social attributes scale (RoSAS)* (In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction).

Chekroun, P., & Brauer, M. (2002). The bystander effect and social control behavior: The effect of the presence of others on people's reactions to norm violations. *European Journal of Social Psychology, 32*(6), 853–867. https://doi.org/10.1002/ejsp.126

Ciardo, F., Lugli, L., Nicoletti, R., Rubichi, S., & Iani, C. (2016). Action-space coding in social contexts. *Scientific Reports, 6*, 22673. https://doi.org/10.1038/srep22673

Ciardo, F., & Wykowska, A. (2018). Response coordination emerges in cooperative but not competitive joint task. *Frontiers in Psychology, 9*, 1919. https://doi.org/10.3389/fpsyg.2018.01919

R Core Team. (2018). R: A language and environment for statistical computing. Vienna. Retrieved from https://www.R-project.org/.

Ciardo, Francesca, Beyer, Frederike, De Tommaso, Davide, & Wykowska, Agnieszka (2020). Attribution of intentional agency towards robots reduces one's own sense of agency. *Cognition, 194*. https://doi.org/10.1016/j.cognition.2019.104109

Czeszumski, A., Ehinger, B. V., Wahn, B., & König, P. (2019). The social situation affects how we process feedback about our actions. *Frontiers in Psychology, 10*, 361. https://doi.org/10.3389/fpsyg.2019.00361

de Bruijn, E. R. A., Miedl, S. F., & Bekkering, H. (2008). Fast responders have blinders on: Erp correlates of response inhibition in competition. *Cortex, 44*(5), 580–586. https://doi.org/10.1016/j.cortex.2007.09.003

de Jong, R., Wierda, M., Mulder, G., & Mulder, L. J. (1988). Use of partial stimulus information in response processing. *Journal of Experimental Psychology: Human Perception and Performance, 14*(4), 682.

Delorme, A., & Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Frith, C. D. (2014). Action, agency and responsibility. *Neuropsychologia, 55*, 137–142. https://doi.org/10.1016/j.neuropsychologia.2013.09.007

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 331–344. https://doi.org/10.1037/0096-1523.14.3.331

Hinz, N.-A., Ciardo, F., & Wykowska, A. (2019). Individual differences in attitude toward robots predict behavior in human-robot interaction. In M. A. Salichs, S. Sam Ge, E.

Barakova, J.-J. Cabibihan, A. R. Wagner, Á. Castro-González, & H. He (Eds.), The 11th International Conference on Social Robotics.
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
Itagaki, S., & Katayama, J. I. (2008). Self-relevant criteria determine the evaluation of outcomes induced by others. *Neuroreport, 19*(3), 383–387. https://doi.org/10.1097/WNR.0b013e3282f556e8
Kestemont, J., Ma, N., Baetens, K., Clément, N., Van Overwalle, F., & Vandekerckhove, M. (2015). Neural correlates of attributing causes to the self, another person and the situation. *Social Cognitive and Affective Neuroscience, 10*(1), 114–121.
Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action. In Psychology of Learning and Motivation. Advances in Research and Theory (Vol. 54, pp. 59–101). Elsevier. doi:https://doi.org/10.1016/B978-0-12-385527-5.00003-6.
Kourtis, D., Knoblich, G., Woźniak, M., & Sebanz, N. (2014). Attention allocation and task representation during joint action planning. *Journal of Cognitive Neuroscience, 26* (10), 2275–2286. https://doi.org/10.1162/jocn_a_00634
Kourtis, D., Sebanz, N., & Knoblich, G. (2010). Favouritism in the motor system: Social interaction modulates action simulation. *Biology Letters, 6*(6), 758–761. https://doi.org/10.1098/rsbl.2010.0478
Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., … Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75–84. https://doi.org/10.1037//1076-898X.8.2.75
Leuthold, H., & Jentzsch, I. (2002). Distinguishing neural sources of movement preparation and execution. *Biological Psychology, 60*(2–3), 173–198. https://doi.org/10.1016/S0301-0511(02)00032-7
Li, P., Jia, S., Feng, T., Liu, Q., Suo, T., & Li, H. (2010). The influence of the diffusion of responsibility effect on outcome evaluations: Electrophysiological evidence from an ERP study. *NeuroImage, 52*(4), 1727–1733. https://doi.org/10.1016/j.neuroimage.2010.04.275
Luck, S. J. (2014). An introduction to the event-related potential technique (second edition). In *A Bradford book*. Cambridge, Massachusetts, London, England: The MIT Press.
Mangun, G. R., Hillyard, S. A., & Luck, S. (1993). Electrocortical substrates of visual selective attention. In D. Meyer & S. Kornblum (Eds.), Attention and performance (XIV, pp. 219–243). Cambridge, Massachusetts: MIT Press.
Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots? *Frontiers in Psychology, 10*, 450. https://doi.org/10.3389/fpsyg.2019.00450
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7
Nomura, T., Sugimoto, K., Syrdal, D. S., & Dautenhahn, K. (2012). Social acceptance of humanoid robots in Japan: A survey for development of the frankenstein syndrome questionnaire. In 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012) (pp. 242–247). doi:https://doi.org/10.1109/HUMANOIDS.2012.6651527.
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869. https://doi.org/10.1155/2011/156869
Perez-Osorio, J., & Wykowska, A. (2019). Adopting the intentional stance toward natural and artificial agents. *Philosophical Psychology, 5*(7), 1–27. https://doi.org/10.1080/09515089.2019.1688778
Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain Cognition, 56*(1), 5–13.
Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006). When things are better or worse than expected: The medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience, 18*(7), 1112–1119. https://doi.org/10.1162/jocn.2006.18.7.1112
Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia, 119*, 302–307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028
Rollin, B. E. (1995). The Frankenstein syndrome: Ethical and social issues in the genetic engineering of animals: Cambridge University Press.
San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience, 6*, 304. https://doi.org/10.3389/fnhum.2012.00304
Schmidt, S., Jo, H.-G., Wittmann, M., & Hinterberger, T. (2016). 'catching the waves' - slow cortical potentials as moderator of voluntary action. *Neuroscience and Biobehavioral Reviews, 68*, 639–650. https://doi.org/10.1016/j.neubiorev.2016.06.023
Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), B11–B21. https://doi.org/10.1016/S0010-0277(03)00043-X
Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: An ERP study of action planning and control in co-acting individuals. *Journal of Cognitive Neuroscience, 18*(5), 859–870. https://doi.org/10.1162/jocn.2006.18.5.859
Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology, 117*, 2341–2356. https://doi.org/10.1016/j.clinph.2006.04.025
Stenzel, A., Chinellato, E., Bou, M. A. T., Del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1073–1077. https://doi.org/10.1037/a0029493
Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., … Teller, A. (2016). *Artificial intelligence and life in 2030: One hundred year study on artificial intelligence: Report of the 2015–2016 study panel*. Stanford: CA. Retrieved from Stanford University website. http://ai100.stanford.edu/2016-report.
Syrdal, D. S., Dautenhahn, K., Kony, K. L., & Walters, M. L. (2009). The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study. In Adaptive and Emergent Behaviour and Complex Systems: Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour.
Syrdal, D. S., Nomura, T., & Dautenhahn, K. (2013). The Frankenstein Syndrome Questionnaire–Results from a quantitative cross-cultural survey. In Proceedings of the International Conference on Social Robotics (pp. 270–279).
Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., … Wahn, B. (2016). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology, 7*, 2039. https://doi.org/10.3389/fpsyg.2016.02039
Vesper, Cordula, Butterfill, Stephen, Knoblich, Günther, & Sebanz, Natalie (2010). A minimal architecture for joint action. *Neural Networks, 23*(8–9), 998–1993. https://doi.org/10.1016/j.neunet.2010.06.002
Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219–232. https://doi.org/10.1177/1745691610369336
Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York.
Zajonc, R. B., Wolosin, R. J., Wolosin, M. A., & Loh, W. D. (1970). Social facilitation and imitation in group risk-taking. *Journal of Experimental Social Psychology, 6*(1), 26–46. https://doi.org/10.1016/0022-1031(70)90074-0