IOP Publishing Journal of Neural Engineering

J. Neural Eng. **15** (2018) 066014 (17pp) https://doi.org/10.1088/1741-2552/aae069

Human-agent co-adaptation using error-related potentials

Stefan K Ehrlich[1] **and Gordon Cheng**

Chair for Cognitive Systems, Department of Electrical and Computer Engineering, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

E-mail: stefan.ehrlich@tum.de

Received 19 May 2018, revised 3 August 2018
Accepted for publication 11 September 2018
Published 28 September 2018

Abstract

Objective. Error-related potentials (ErrP) have been proposed as an intuitive feedback signal decoded from the ongoing electroencephalogram (EEG) of a human observer for improving human-robot interaction (HRI). While recent demonstrations of this approach have successfully studied the use of ErrPs as a teaching signal for robot skill learning, so far, no efforts have been made towards HRI scenarios where mutual adaptations between human and robot are expected or required. These are collaborative or social interactive scenarios without predefined dominancy of the human partner and robots being perceived as intentional agents. Here we explore the usability of ErrPs as a feedback signal from the human for mediating co-adaptation in human-robot interaction. *Approach.* We experimentally demonstrate ErrPs-based mediation of co-adaptation in a human-robot interaction study where successful interaction depended on co-adaptive convergence to a consensus between them. While subjects adapted to the robot by reflecting upon its behavior, the robot adapted its behavior based on ErrPs decoded online from the human partner's ongoing EEG. *Main results.* ErrPs were decoded online in single trial with an avg. accuracy of 81.8% $\pm$ 8.0% across 13 subjects, which was sufficient for effective adaptation of robot behavior. Successful co-adaptation was demonstrated by significant improvements in human-robot interaction efficacy and efficiency, and by the robot behavior that emerged during co-adaptation. These results indicate the potential of ErrPs as a useful feedback signal for mediating co-adaptation in human-robot interaction as demonstrated in a practical example. *Significance.* As robots become more widely embedded in society, methods for aligning them to human expectations and conventions will become increasingly important in the future. In this quest, ErrPs may constitute a promising complementary feedback signal for guiding adaptations towards human preferences. In this paper we extended previous research to less constrained HRI scenarios where mutual adaptations between human and robot are expected or required.

Keywords: electroencephalography (EEG), brain–computer interface (BCI), event-related potentials (ERP), error-related potentials (ErrP), error monitoring, human-robot interaction (HRI), co-adaptation

S Supplementary material for this article is available online

[1] Author to whom any correspondence should be addressed.

1. Introduction

Over the last two decades, research on non-invasive brain–computer interfaces (BCI) (Wolpaw *et al* 2002) has gained increased interest in error-related potentials (ErrPs). ErrPs are event-related potentials (ERP) (Blankertz *et al* 2011) occurring in response to the human recognition of both self-inflicted (Miltner *et al* 1997, Falkenstein *et al* 2000, Botvinick *et al* 2001, Holroyd and Coles 2002) and/or observed (van Schie *et al* 2004) erroneous actions. The underlying neural process is understood to be related to error-/performance monitoring in the brain, crucial for goal-directed behavior, decision making, error handling as well as adaptation and learning (Ridderinkhof *et al* 2004, Garrido *et al* 2009, Alexander and Brown 2011, Ullsperger *et al* 2014). ErrPs are a reliable effect observable in the human electroencephalogram (EEG) and their decoding from EEG signals has repeatedly shown to be robust across recording sessions (Chavarriaga and Millán 2010) and high-performant with single trial classification accuracies around 70%–80% (Ferrez and Millán 2005, 2008a, Chavarriaga *et al* 2014). Schalk *et al* (2000) were among the first to propose the use of ErrPs for online improvements of BCI decoders. They demonstrated that ErrPs occur in response to the subject's observation of the BCI delivering wrong output, e.g. mismatching the subject's intended command the BCI was ought to execute. This discovery led to a series of studies simulating and demonstrating the efficacy of simultaneous ErrP-decoding for online adaptation of BCI decoders, in particular for sensorimotor BCIs (Blankertz *et al* 2003, Ferrez and Millán 2008b), but also P300-based speller BCIs (Schmidt *et al* 2012, Spüler *et al* 2012a, 2012b).

More recently, ErrPs have been proposed as a feedback signal from the human for guided adaptations of physical robotic systems (Iturrate *et al* 2010, 2015, Kreilinger *et al* 2012, Ehrlich and Cheng 2016, Kim *et al* 2017, Salazar-Gomez *et al* 2017, Welke *et al* 2017). The basic concept is to harvest ErrP responses from a human observer upon recognition of erroneous or inappropriate robot actions in order to adapt or improve the robotic system post-hoc or on-the-fly. This approach is particularly promising as a complementary method for validating and improving robotic systems and human-robot interaction (HRI), because: (1) ErrPs are naïve responses which require no mental effort from the human observer. (2) ErrPs can be decoded in real-time, allowing for online adaptations of the robotic device without interruption of ongoing interaction with the human partner. (3) ErrPs are understood to be sensitive to violations of expectations (Oliveira *et al* 2007, Sallet *et al* 2007) and as such comprise an implicit and immediate feedback, informative (3a) for improving the robotic system to better align with the observer's expectation, and (3b) possibly informative with regard to the observer's overall assessment of the robotic system and/or the quality of interaction. Recent works have successfully demonstrated the use of online decoded ErrPs from a human observer for intuitive reinforcement learning (RL) of robot skills, e.g. execution of trajectories in an end effector reaching task (Iturrate *et al* 2015), association of objects in a sorting task (Salazar-Gomez *et al* 2017), as well as recognition and imitation of human gestures (Kim *et al* 2017). While these works showed promising results of this highly innovative approach, they were primarily concentrated on using ErrPs as a teaching signal for robot skill learning. A question that remains open is whether this ErrP-based feedback signal can also be useful in situations where both human and robot are required to adapt to each other to converge to a consensus in the given joint task. That is, situations in which there is no explicit 'right' behavior (policy) the robot is supposed to be taught, but the human partner may as well adapt to the robot. Approaching this question is important regarding human interaction with systems that have a form of intentional agency, e.g. HRI in the context of collaborative or social interactive scenarios. This contrasts to human interaction with robotic systems that are primarily used as tools supposed to fulfill an explicit function (explicit 'right' policy), e.g. a neural prosthesis.

The current study explored the usability of ErrPs as a feedback signal in the context of human-agent co-adaptation. The approach used is schematically described in Figure 1 and conceptually assumes the interaction between two partners: (1) An intentional artificial agent with a policy π determining its behavior based on a set of behavioral states S, actions A, and goals/intentions G. (2) A human partner, interacting with that agent based on a belief of the agent's policy π'. While the agent is provided feedback through online decoded ErrPs to gradually adapt its policy π to the human's belief π', the human partner may gradually adapt his/her belief π' to the agent's policy π by reflecting upon its behavior. As such, both systems (human and agent) are adaptive, allowing for mutual adaptation with the aim to converge to a consensus in form of an alignment of the human's belief and the agent's actual policy: $\pi'_{final} \approx \pi_{final}$.

The conceptual approach was implemented in form of a human-robot social interactive repeated guessing game where the human partner has to guess, from a humanoid robot's gazing behavior, which of three available objects was selected by the robot. While the human's task was to learn to infer the robot's intentions/goals by observing and interpreting the robot's gazing behavior, the robot's task was to learn to convey its intentions/goals via gazing behavior to the human partner; efficient interaction required their convergence to a consensus by co-adaptive learning of both parties. We experimentally demonstrate that ErrPs decoded online from the ongoing EEG of the human partner can successfully be used to mediate and establish co-adaptation between human and robot as indicated by significant improvements in interaction performance.

With this extended perspective we aim to make the following contributions in line with ongoing research on the use of ErrPs for HRI:

- We demonstrate the usability of ErrPs for mediating co-adaptation in HRI. This relaxation of interaction constraints—permitting mutual adaptation—is particularly important with regard to HRI scenarios where the human partner does not have a predefined dominant role (principal or teacher role). Scenarios, in which adaptations of the human to the robot are expected or even necessary for successful interaction, such as in collaborative or social interactive HRI.

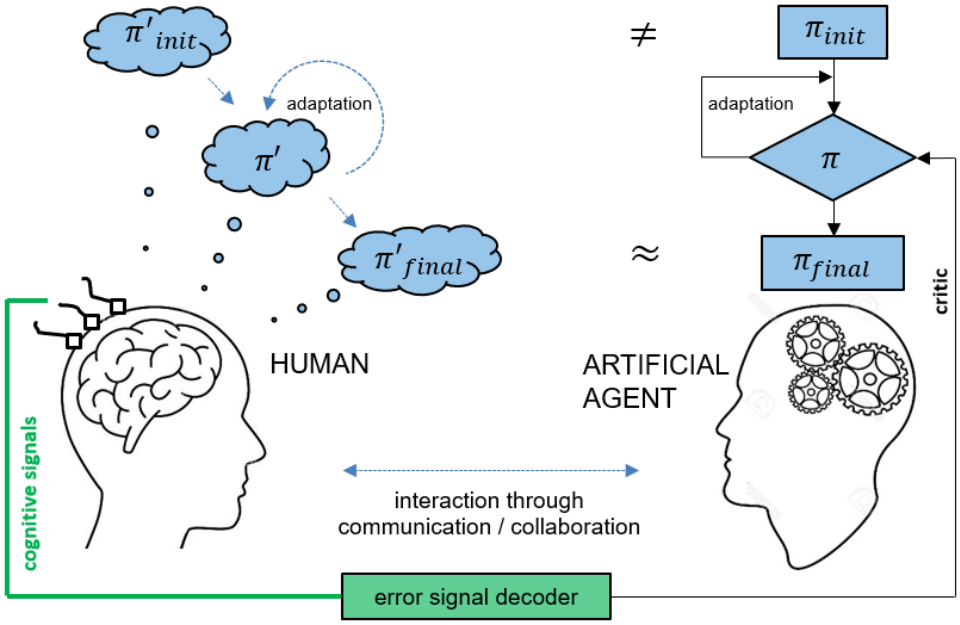

Figure 1. Conceptual approach: in interaction with an agent, the human holds a mental model (belief) of the agent's policy π' to predict its future behavior, which can be based on prior expectations π'_{init} and is further adapted during interaction. ErrPs, online decoded from neural activity of the human partner are, provide as a critic for guided adaptation of the agent's actual model π. This creates a two-party co-adaptive system allowing both human and agent seeking consensus in form of an alignment of the human's belief and the agent's actual policy: $\pi'_{final} \approx \pi_{final}$.

- Previous works adapted the robot's behavior using ErrP feedback based on single, explicitely erroneous robot actions (Iturrate *et al* 2015, Salazar-Gomez *et al* 2017, Kim *et al* 2017). In more complex robot behavior, however, individual robot actions are more likely to occur in rapid succession and not to be temporally well isolated, the latter being a prerequisite for reliable ErrP decoding. Here we demonstrate robot adaptation based on ErrPs arising from and reflecting the human's interpretation of the robot's intention/goal, with the latter comprising a sequence of actions instead of a single occurrence. Along this line, we propose and successfully employ an ErrP-based episode update strategy with delayed reward for online adaptation of the past sequence of robot actions.

The paper is structured as follows: in the subsequent sections the experimental paradigm (section 2.1), design and tasks (section 2.2) are described in detail, followed by a thorough description of the implementation of the technical components of our approach (EEG-based online decoding of ErrPs and corresponding online adaptation of robot behavior) in section 2.3. The main results of efficacy of online ErrP decoding and human-robot co-adaptation are reported in section 3, followed by a discussion of the results in light with the outlined contributions of this paper in section 4. Section 5 concludes the paper.

2. Methods

2.1. Experimental paradigm

The experimental paradigm is schematically described in figure 2(a). Three objects were located in between subject and robot. The robot would select one among the three objects (unknown to the subject) which denoted its covert goal/intention g and subsequently started executing a gaze pattern (action sequence) by turning its head (actions: A) towards the objects and the subject (states: S). The subject's task was to guess the robot's initially selected object from observing its gaze behavior: The subject may for instance consider which object the robot fixated more often or for the longer duration. Eventually, the robot would reveal the actual object it has initially selected, resulting in the subject experiencing a match or mismatch with the object he/she believed the robot had selected. In that moment the subject's ongoing EEG signals would be classified into an *error-* (mismatch) or *non-error* (match) response and used as a negative or positive reward for adaptations of the robot's gaze behavior policy π. Meanwhile, the subject may update his/her prior belief π' about the robot's gaze behavior to improve guessing in subsequent games. We hypothesize that by using the ErrP feedback for iterative robot adaptation would eventually converge to robot gazing behavior which facilitates the subject to correctly infer the robot's selected object. To what extend this convergence is driven by the human adapting to the robot or the robot adapting to preconceptions of the human is deliberately kept flexible to investigate the feasibility of ErrP-based mediation of co-adaptation as outlined in the introduction.

2.2. Experimental design

2.2.1. Participants. Eighteen healthy subjects participated in the study. The data of the first two subjects were discarded due to technical problems during the experiment. The remaining sixteen subjects were 7 females, average age: 29.2 $\pm$ 5.0, and

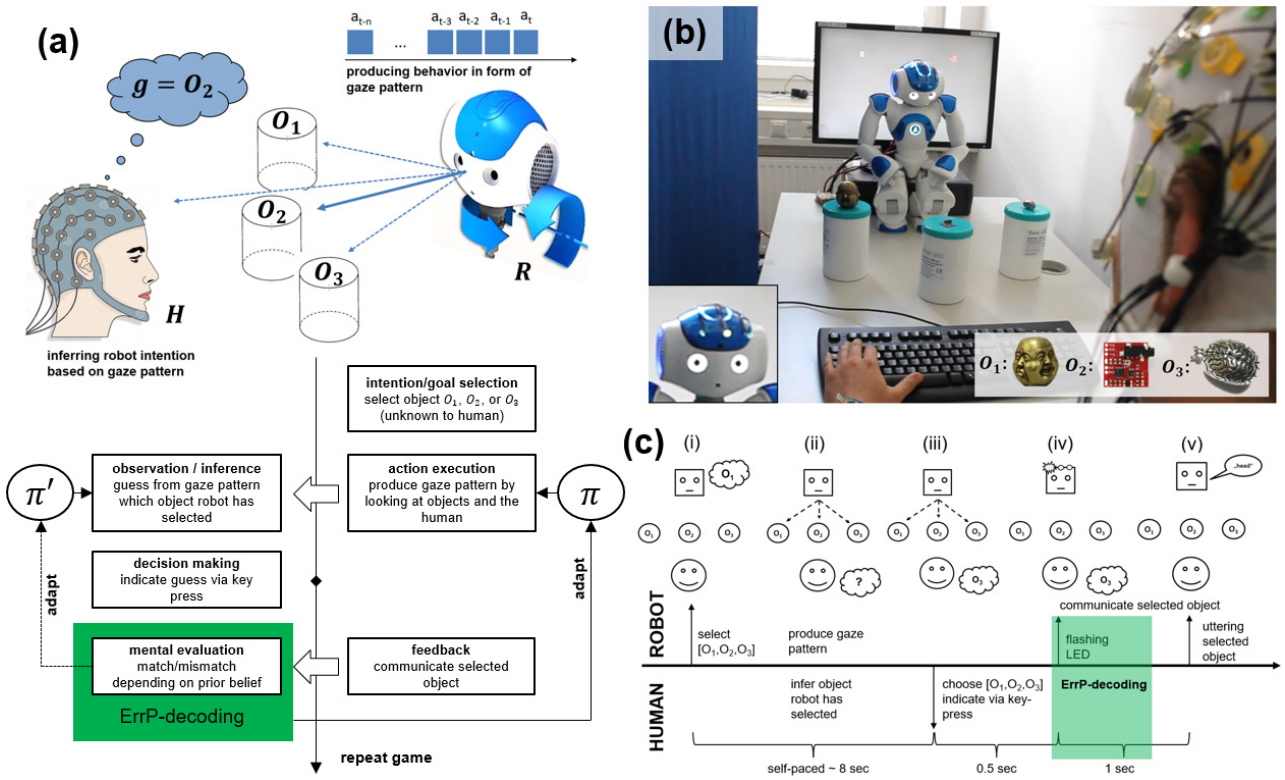

Figure 2. (a) Experimental paradigm: human subject and robot play a guessing game in which the robot covertly selects one out of three objects. Subsequently the robot produces a gaze pattern based on which the subject has to guess the secret object. The subject's brain responses are measured (marked in green) and used as a feedback signal to adapt the robot's gaze behavior policy π, while the subject may adapt the prior belief π' about the robot's gaze behavior policy. (b) Experimental setup from the perspective of a subject. (c) Trial structure of a single guessing game with the corresponding moment of ErrP decoding marked in green.

all right-handed. The study was approved by the institutional ethics review board of the Technical University of Munich. All subjects were informed about the tests before its conduction. They participated voluntarily and gave written consent. Participants were paid an honorarium of 8 EUR/h for their participation.

2.2.2. Experimental tasks. An overview of the experimental tasks is provided in table 1. Each experiment consisted of two parts conducted in the following order: (1) open-loop calibration session (CALIB), (2) four closed-loop co-adaptation sessions (CORL). In both parts, the subject was asked to repeatedly play the guessing game together with the robot (for the remainder of this paper, a single game is also called trial; for technical details about the trial structure the reader is referred to section 2.3.1 and figure 2(c)). The first part of the experiment (CALIB) had the purpose of collecting EEG data for calibrating subject-specific ErrP-decoders utilized afterwards for online ErrP-decoding during the co-adaptation sessions (CORL). During CALIB, the robot's gaze behavior policy was pre-programmed and not adapting; during CORL, the robot's gaze behavior policy was online adapted based on the ErrPs decoded from the subject's EEG signals while interacting with the robot. We opted for an experimental design integrating calibration and online application in a single recording session per subject to avoid the possibility of day-to-day data variability. Furthermore, design choices on the number of trials for CALIB and CORL were made to keep the duration of the experiment around 1 h maximum to avoid subjects suffering from concentration lapses resulting in data quality degradation.

2.2.2.1. Open-loop calibration session (CALIB). During this experimental task, the robot's gaze behavior followed a deterministic behavior in which the robot tended to look at the selected object more often or remained gazing at it (for details about the technical implementation the reader is referred to section 2.3.2). With this gaze behavior, subjects achieved high accuracies in guessing the robot's selected object (>95%). Participants performed in total 150 guessing games (trials) during the calibration session in three blocks of 50 trials each, resulting in a total duration of 20.4 ± 6.7 min. Pilot experiments showed that subjects could perform 50 trials in a row without reporting notable concentration drops. The breaks in between blocks allowed the subjects to relax and prepare for the next block. To control the number of error events, false feedback was introduced with a probability of $p_{err} = 0.3$. Usually, false feedback rates are chosen around 20% (Ferrez and Millán 2005, Chavarriaga *et al* 2014, Iturrate *et al* 2015). Here, an increased false feedback rate was used to obtain a higher number of error observations given the limited number of 150 trials. False feedback was realized as wrong robot feedback irrespective of whether the subject guessed correctly. This resulted in approximately 35% error (mismatch) events combining false feedback and subject guessing mistakes. The EEG data recorded during the calibration session

Table 1. Overview of the experimental tasks and corresponding purpose in the order of conduction, covering part 1 (open-loop calibration session—CALIB), and part 2 (four separate closed-loop co-adaptation sessions—CORL-I,-II,-III, and -IV). The duration of the entire experiment was approximately 1 h per subject, including breaks.

Part	Short name	Short description and purpose	Duration
1 Open-loop calibration session	CALIB	***Task:*** To guess the robot's selected object, indicating guess via key-press. ***# trials:*** 150 (3 blocks of 50 trials each) ***Purpose:*** Collect EEG data for subsequent calibration of subject-specific ErrP-decoders ***Robot gaze policy:*** pre-programmed and non-adaptive. Elicitation of ErrPs with random occurrences of false-feedback events with a probability of $p_{err} = 0.3$.	15–25 min
	ErrP decoder calibration	Automatic calibration of subject-specific ErrP decoder based on data collected during CALIB	5 min
2 Closed-loop co-adaptation sessions	CORL-I	***Task:*** To guess the robot's selected object, indicating guess via key-press. ***# trials:*** 50 ***Purpose:*** Online application of ErrP decoder for mediating human-robot co-adaptation ***Robot gaze policy:*** Initial uniformly random gaze behavior; updated after each trial based on the classified outcome of the corresponding online decoded ErrP	6–8 min
	CORL-II	Same as CORL-I with reinitialization of gaze policy	6–8 min
	CORL-III	***Task:*** To guess the robot's selected object *without* overtly indicating guesses via key-press (compared to CORL-I,-II, and -IV). Robot performed gaze behavior for a pre-defined fixed duration. ***# trials:*** 50 ***Purpose:*** Online application of ErrP decoder for mediating human-robot co-adaptation *without explicit decisions* from the human partner (sole observation and mental reflection upon the robot's gaze behavior) ***Robot gaze policy:*** same as in CORL-I,-II, and -IV with reinitialization of gaze policy	6–8 min
	CORL-IV	Same as CORL-I with reinitialization of gaze policy	6–8 min

was afterwards used to build a subject-specific ErrP-decoder subsequently to be employed for online ErrP-decoding during the co-adaptation sessions. During CALIB, subjects indicated their guess by keypress and received feedback about their correct and wrong number of guesses displayed on a computer screen. Since the calibration session had a comparably long duration and was rather monotonous (non-adaptive robot behavior), this feedback was introduced as a means for self-monitoring to encourage subjects to improve and maintain their performance throughout the session.

2.2.2.2. Closed-loop co-adaptation sessions (CORL). After the calibration session, each participant performed four separate online co-adaptation runs, each consisting of 50 trials, with an average duration of approx. 6–8 min per run (corresponding to the duration of one block during CALIB). We opted for the conduction of several independent runs per subject, as co-adaptation may fail due to uncontrollable and random factors (e.g. stochastic sampling of robot actions, self-paced subject response; further details are provided in section 2.3). Furthermore, a single CORL was kept below 10 min to restrict subject frustration if co-learning would turn out unsuccessful. In the beginning of each run the robot gaze behavior policy was reinitialized such that gazing was uniformly random distributed among the three objects and therefore allowed for no informed guesses of the selected object (chance-level $p = 1/3$). In each trial, the subject's ErrP response to the robot revealing its initially selected object was decoded online from the ongoing EEG signals and utilized to update the robot's gazing behavior policy. In three of the four runs (CORL-I, CORL-II, CORL-IV), the procedure was identical to the open-loop calibration session, in that subjects indicated their guess by keypress. In CORL-III, subjects were asked to observe the robot's behavior for a pre-defined time only, without communicating their guesses via keypress. We introduced this additional run to investigate whether explicit actions linked to the decisions (as in the form of keypresses) of the subject were required or whether covert beliefs/decisions (sole mental reflection without explicit actions) are sufficient for successful co-adaptation. As such, CORL-III served as a preliminary usability test of the proposed ErrP-based approach in a more naturalistic and less constrained setting, where no explicit information from the subject is available. Unlike in CALIB, in CORL, no feedback on the number of correct guesses was provided to subjects. This had no notable effect on the observability and decodeability of ErrPs (see section 3.1, and figure 3(b)).

2.2.3. Instructions. Subjects were asked to guess the robot's selected object by inferring information from its gazing behavior. They were, however, not given any specific hints on what to focus in particular. After the calibration session, subjects were furthermore informed that during the online co-adaptation runs the robot's gaze behavior may change as it is

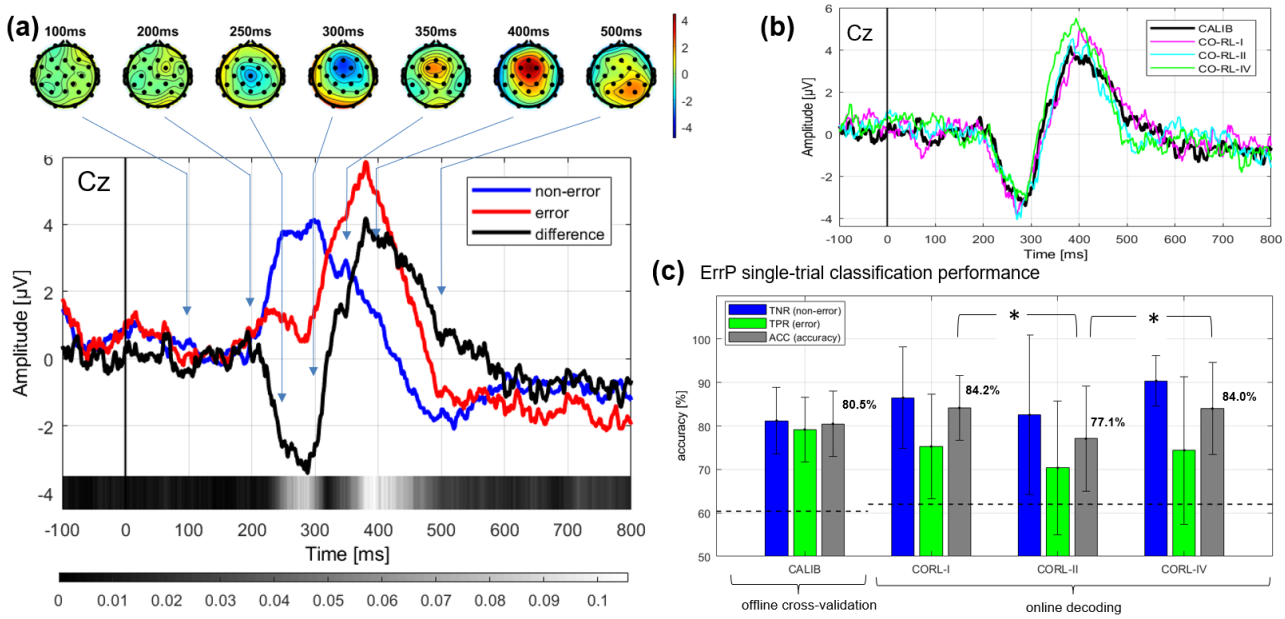

Figure 3. (a) Grand average ($n = 13$) ERP time-courses over channel Cz time-locked to the onset of feedback presentation for each class of events (blue: non-error, red: error) and the difference grand average (black: error minus non-error). The r^2-values for between non-error and machine-error are depicted below the plot, with brighter colors indicating higher values. The difference grand average is furthermore depicted as topographic plots for the main peaks above each plot and in form of a spatio-temporal activity matrix across all channels and specific time points above the plot (Figure style adopted from Iturrate *et al* (2013).) (b) Comparison of the difference grand average across the calibration session and the co-adaptation runs I, II, and IV. The comparison shows high resemblance of the difference ERPs across experimental sessions. (c) ErrP single trial classification performance for offline cross-validation based on the calibration data and online decoding performance during the co-adaptation runs. The black dashed lines indicate the theoretical chance-level of 60.37% for CALIB and 62.0% for individual CORLs. For CALIB, chance-level was exceeded in all subjects; for CORL, chance-level was exceeded in all but three cases (s11/CORL-II, s14/CORL-II, s06/CORL-IV). Online decoding accuracies were significantly different between CORL-I and CORL-II and between CORL-II and CORL-IV, indicated by the asterisks.

subject to adaptations based on their ongoing brain activity. No further details about the implementation of the experiment were provided.

2.3. Systems: human, agent, brain-machine interface

2.3.1. Systems overview. Figure 2(b) shows the experimental setup from the perspective of a subject sitting approximately 150 cm in front of a humanoid robot. The robotic platform chosen for the experiment was the humanoid robot NAO. NAO is a commercially available (SoftBank Robotics) 58 cm tall humanoid robot with 21–25 degrees of freedom (Gouaillier *et al* 2008) that was controlled in this experiment by a program running on an external PC connected to the robot via local area network (LAN). The only body part of the robot used was the head with pitch- and yaw-movements. Three arbitrarily chosen physical objects were located on top of cylindrical containers in fixed positions between the subject and the robot (O_1: a metal Buddha head—left, O_2: an electric circuit board—middle, O_3: a metal brain keychains—right). The robot's forehead was equipped with three identical green light emitting diodes (LED) geometrically aligned with the three objects and placed in a distance of approx. 1.5 cm from each other (see figure 2(b), bottom left corner). The LEDs were controlled via a 4-channel digital analog converter (Phidget). We used the LEDs in the experiment as the visual feedback to communicate the robot's initially selected object to the subject (figure 2(a): 'feedback') with a fixed representation (O_1: left LED, O_2: middle LED, O_3: right LED). Subsequently, subjects received an additional auditory feedback in form of the robot speaking out the name of the chosen object (O_1: 'The head', O_2: 'The circuit', O_3: 'The brain'). The robot's choice was communicated in two ways for the following reasons: LED-based feedback was introduced because of its high saliency and perceptual simplicity, expected to result in more distinct brain-responses in contrast to perceptually complex or gradually unfolding stimuli which have been reported to result in attenuated ErrP responses (Omedes *et al* 2015, Ehrlich and Cheng, 2016, Welke *et al* 2017). The subsequent additional robot speech feedback was introduced to increase subject's engagement in the experiment, as robot talking has been reported to foster engagement in HRI (Sidner *et al* 2004). Subjects were instructed to particularly attend the LED feedback. Behind the robot, a computer screen was located which served to provide the subject with additional information about the number of correct (left, green) versus incorrect (right, red) guesses during the calibration session (CALIB). During the co-adaptation runs (CORL), this feedback was not provided. Participant responses (figure 2(a): 'decision making') were performed with the left hand and registered with the following keys of an ordinary computer keyboard (O_1: key '1', O_2: key '2', O_3: key '3'). The experiment was realized in a single program using the Python-based NAOqi-library (robot control), the Phidgets-library (LED control), Psychopy library (keyboard and screen control) (Peirce 2007) and executed on an Intel®Core™ i5 CPU 750@2.67 GHz. Furthermore, this

program received input from the ErrP decoder during the co-adaptation runs which was executed on a different PC via the TCP/IP-based 'labstreaminglayer' protocol (Kothe 2014).

2.3.1.1. Trial (guessing game) structure. Figure 2(c) shows the structure of a single trial: (a) the trial started with the robot gazing at the human, and performing a uniform random selection of either one of the objects $g \in G$, as well as a uniform random selection of an initial gaze state $s_{init} \in S$ and (b) subsequently started alternatingly gazing in a fixed pace at the objects and the subject based on the current policy π, starting from the initial state s_{init}. Meanwhile, the subject's task was to guess the robot's choice from its gaze behavior and (c) indicate the guess with a corresponding left-hand key-press in a self-paced fashion. Upon keypress response, the robot stopped action execution and turned its head back at the subject. (d) After a delay of half a second (for avoiding any superposition of event-related brain activity in response to the preceding robot head movement), the robot announced the selected object to the subject by lightening up the corresponding LED attached to its head. This stimulus was delivered for 1 s, during which the ErrP response was decoded (the duration of a typical ErrP is typically no longer than 600–800 ms (Chavarriaga *et al* 2014)) and (e) afterwards, the additional auditory robot speech feedback was provided to the subject. This trial structure applied to CALIB and CORL-I, -II, and -IV. In CORL-III subjects were not required to indicate their guesses via key-press responses (no explicit decision making), ending the robot's action execution phase, but only to observe the robot's behavior. In CORL-III, the duration of robot action execution was therefore fixed to 15 gaze transitions. This parameter was determined empirically based on the average number of gaze transitions until subject decision during the calibration sessions of a series of pilot experiments. Other than step (b) and (c), no other parts of the trial were affected by the modifications in CORL-III.

2.3.2. Robot gaze policy and behavior.

2.3.2.1. Intention/goal selection. The possible selections the robot can choose from defined the agent's set of goals/intentions $G = \{g_{O1}, g_{O2}, g_{O3}\}$. The robot's selection always followed a uniform random choice among the three options.

2.3.2.2. Gaze policy. For realizing gaze behavior in robotic systems, earlier works have proposed the use of probabilistic state machines, e.g. for establishing joint attention (Lanillos *et al* 2015). On this basis, the robot's internal gaze policy was realized as a discrete state-space model with four states, $S_\pi = \{s_{objInt}, s_{othObjx}, s_{othObjy}, s_{human}\}$, with s_{objInt}: gazing at selected object; $s_{othObjx}$, $s_{othObjy}$: gazing at one of the other objects; and s_{human}: gazing at human. An action is considered a transition from one gaze state to another or remaining in the current gaze state, leading to 16 possible state-action pairs: $A_{i,j},\ i,j = S_\pi$. The policy $\pi(a_i|s_j) \in [0,1] \subset \mathbb{R}$ determined the gaze behavior described by the probability of taking action a_i in state s_j (gaze transition from state s_j to s_i). The decision for the next action was always performed by a weighted random selection among the four possible actions in the current state (remain in current state or transit to one of the three other states). This way of realizing action selection implicitely introduced an exploration-exploitation behavior, an approach used to foster successful policy convergence (Sutton and Barto 1998). In our case, equiprobable distributions among the set of actions to be selected from, drive exploratory behavior and divergent probabilities drive exploitation behavior.

2.3.2.3. Action (gaze pattern) execution. The robot gaze behavior resulted from a fixed mapping between the covert policy-states S_π and the overt action-execution states S_{act} with corresponding pre-defined robot head angles, pitch ψ and yaw Θ: $S_{act} = \{s_{O1}, s_{O2}, s_{O3}, s_H\}$, with s_{O1}: gazing at O_1, $\psi_{O1} = 25°$, $\Theta_{O1} = -20°$; s_{O2}: gazing at O_2, $\psi_{O2} = 25°$, $\Theta_{O2} = 0°$; s_{O3}: gazing at O_3, $\psi_{O3} = 25°$, $\Theta_{O3} = 20°$; and s_H: gazing at subject, $\psi_H = 0°$, $\Theta_H = 0°$. The mapping depended on the selected object and was realized as follows:
$S_{\pi\to act}(g_{O1}) = \{s_{objInt\to O1}, s_{othobjx\to O3}, s_{othobjy\to O2}, s_{human\to H}\}$,
$S_{\pi\to act}(g_{O2}) = \{s_{objInt\to O2}, s_{othobjx\to O3}, s_{othobjy\to O1}, s_{human\to H}\}$,
$S_{\pi\to act}(g_{O3}) = \{s_{objInt\to O3}, s_{othobjx\to O1}, s_{othobjy\to O2}, s_{human\to H}\}$.
During action execution, the robot performed one action (state-transition) per 400 ms (2.5 Hz); each gaze shift was executed with a fixed speed of 15% and 10% of the maximum joint speed for pitch- and yaw-movements, respectively. These parameters were manually tuned such that the frequency of gaze shifts was maximized while preserving smooth, non-jerky gazing behavior and approximately conforming to human timing of head posture shifts during conversation (Hadar *et al* 1984).

2.3.2.4. CALIB gaze policy. During the calibration session, the gaze policy was pre-programmed with high probabilities for the following state-action pairs, such that the probabilities of all four possible actions in each state summed up to one: $p(a_{objInt}|s_{human}) = 0.85$, $p(a_{objInt}|s_{othObjx}) = 0.85$, $p(a_{objInt}|s_{othObjy}) = 0.85$, $p(a_{objInt}|s_{ObjInt}) = 0.85$ and low probabilities of $p = 0.05$ for all remaining state-action pairs. This resulted in gaze behavior in which the robot tended to fixate the selected object or gaze at it more often.

2.3.2.5. CORL gaze policy initialization. At the beginning of each co-adaptation run, all state-action pairs of the robot's gaze policy were initialized with $p(a_i|s_j) = 0.25$. This resulted in uniform random gaze behavior which allowed no informed guesses about the selected object (chance-level $p = \frac{1}{3}$). As the co-adaptation runs proceeded, the gaze policy underwent one update per trial based on the outcome of online decoded ErrPs. The procedures for ErrP decoding and gaze policy update are described in detail in the following sections.

2.3.3. Decoding of ErrPs.

2.3.3.1. EEG data recording. In all parts of the experiment, EEG data were acquired with a Brain Products actiChamp amplifier equipped with 32 active EEG electrodes arranged according to an extended international 10–20 system (Homan *et al* 1987) (FP1, FP2, F3, F4, F7, F8, FC1, FC2, FC5, FC6, C3, C4, T7, T8, CP5, CP6, P3, P4, P7, P8, TP9, TP10, O1, O2, Fz, Cz, Pz, EOG1, EOG2, EOG3). All leads were referenced to the average of TP9 and TP10 (average mastoids

referencing) and the sampling rate was set to 1024 Hz. The impedance levels of all leads were kept below 10 kΩ. Three channels were used for capturing electrooculogram (EOG1-3) signals in three locations of the participant's face (forehead, left and right outer canthi) according to a method suggested by Schlögl *et al* (2007). The data was transferred via USB to a separate recording PC (Intel®Core™ i5 CPU 750@2.67 GHz). Data recording, pre-processing, and ErrP decoding was performed using OpenViBe (Renard *et al* 2010) together with customized processing functions implemented in MATLAB®.

2.3.3.2. Offline modeling of ErrP-decoder. For each subject, an individual ErrP-decoder was trained based on the data collected during the calibration session. The following procedure was implemented such that an ErrP-decoder was automatically trained right after the calibration session to be employable for online ErrP-decoding during the subsequent co-adaptation runs (this procedure took about 5 min): The data was first filtered with a causal first-order Butterworth FIR bandpass filter with cutoff frequencies 0.5 and 20 Hz. Then, EOG activity (horizontal and vertical) was reduced in the data by using a regression method proposed by Schlögl and colleagues (Schlögl *et al* 2007). The data was then re-referenced to common average. The data was further segmented into data epochs [0,1] sec for non-error- and error-events time-locked to the moment of presentation of LED feedback. Data segments in which the subject's guess did not match the feedback of the robot were labeled as error-events, with no distinction of whether the mismatch resulted from the human incorrect guess or the robot's false feedback; data segments in which the guess matched the feedback were labeled as non-error-events. All data segments were then normalized by subtracting their individual means for each channel/segment. In the context of single-trial classification of ErrPs, temporal features extracted from the time series have been shown to lead to high classification performances and mostly outperformed other types of features (Iturrate *et al* 2010, 2015, Ehrlich and Cheng 2016). Therefore, temporal features were used in this work: The arithmetic mean of the signal amplitude in pre-defined windows relative to the moment of feedback presentation was computed, such that all relevant components of the ErrP event-related potential were covered (windows: 150–250 ms, 200–300 ms, 250–350 ms, 300–400 ms, 350–450 ms, 400–500 ms, 450–550 ms), resulting in a total of 189 temporal features per epoch (27 channels × 14 windows). The features were then used to train a regularized version of the linear discriminant analysis classifier (rLDA) (Friedman 1989). The rLDA classifier has been established as a robust method to discriminate mental states based on EEG signals in the field of BCI (Blankertz *et al* 2011). The LDA discriminant function is the hyperplane discriminating the feature space corresponding to two classes: $y(x) = \text{sign}(w^T x + b)$, with x being the feature vector, w being the normal vector to the hyperplane (or weight vector), b the corresponding bias, and $y(x) \in \{-1, 1\}$ the classifier decision. The weight vector and bias were computed by $w = (\widehat{\mu}_2 - \widehat{\mu}_1)(\widetilde{\Sigma}_1 + \widetilde{\Sigma}_2)^{-1}$ and $b = -w^T(\widehat{\mu}_1 + \widehat{\mu}_2)$, with $\widehat{\mu}_j$ being the class-wise sample means, and $\widetilde{\Sigma}_j$ the class-wise regularized covariance matrices. Regularization aims at minimizing the covariance estimation error by penalizing very small and large eigenvalues. This leads to robust covariance estimates even for high dimensional feature spaces (Blankertz *et al* 2011) as in our case. The regularized covariance matrices were computed by $\widetilde{\Sigma}_j = (1-\lambda)\,\Sigma_j + \lambda I$, with $\lambda \in [0, 1] \subset \mathbb{R}$ being the shrinkage parameter and I the identity matrix (Schäfer and Strimmer 2005). The optimal shrinkage parameter was determined using 10-times-10-fold cross-validation based grid search for $\lambda = [0, 1]$ in steps of 0.05. To avoid the classifier favoring one class over the other, each time and fold, the number of trials per class was balanced by random pick and replace (please note that the number of trials per class was initially unbalanced with ~65% non-error and ~35% error trials). The λ with the highest cumulative accuracy of non-error (true-negative rate, TNR) and error (true-positive rate, TPR) recognition was selected and used to train the final rLDA classifier based on all trials of the calibration data. Also, in this final step, the numbers of trials per class were balanced by random pick and replace. To increase the likelihood that most of the calibration trials were used at least once, this procedure was repeated 1000 times and the weights w and bias b of individual models were averaged to obtain a single final rLDA classification model. The chance-level threshold for this binary classification problem is 60.37% for both TNR and TPR, given balanced number of trials for decoder calibration (inverse cumulative binomial distribution with number of trials $nTrials = 150 * p_{err}$ and $p_{err} = 0.35$; probability of success $p_{success} = 0.5$; confidence threshold $p = 0.05$). Subjects in which either TNR or TPR did not exceed the above threshold were excluded from all further analyses.

2.3.3.3. Online ErrP decoding. During the co-adaptation runs, the ErrP-decoder trained based on the calibration data was used to decode ErrPs in the ongoing EEG acquired from the subject during interaction with the robot. The signals were continuously bandpass filtered using a causal first-order Butterworth FIR bandpass filter with cutoff frequencies 0.5 and 20 Hz (identical filter parameters as used during offline modeling). EOG activity was continuously reduced by applying the EOG decorrelation matrix obtained from the calibration data. Finally, the continuous signals were re-referenced to common average. Upon occurrence of a feedback event (robot communicating decision via flashing one out of three LEDs), the respective data segment (time-locked to the event) was processed in the same way as in offline modeling: (1) single-trial normalization, (2) temporal features extraction, and (3) classification into non-error or error event using the rLDA classifier trained on the calibration data.

2.3.4. ErrP-based agent policy adaptation. For the ErrP-based policy adaptation we decided to employ a learning paradigm based on policy gradient methods, a subform of RL (Sutton and Barto 1998). Among others, the main advantages of policy gradient methods compared to more sophisticated RL methods, such as Q-learning are as follows: first, policy gradient methods act on-policy directly which facilitates the interpretation of policy adaptations in contrast to value function based approaches. This property comes in handy for the

qualitative analysis of emergence of gaze behavior (see section 3.3). Second, policy gradient learning has been proposed as the method of choice for RL for humanoid robots as they can deal with complex learning tasks involving many degrees of freedom (Gullapalli *et al* 1994, Peters and Schaal 2008). Although not in the focus of this paper, this property favors the generalizability and scalablity of our approach to more complex robot behavior and interaction scenarios. The policy update function is given in equation (1) which was executed at the end of each trial during the co-adaptation runs, starting with the initial policy π_{init} with equal probabilities for all state-action pairs $p(a_i|s_j) = 0.25$. In short, for the computation of the parameters of the new policy π^{t+1}, to be employed in the next trial, the parameters of the old policy π^t were merged in a weighted fashion with the empirical distribution of the observed state-action pairs during the current trial:

$$\pi^{t+1}(a_i|s_j) = \pi^t(a_i|s_j) + \alpha R \sum_{k=1}^{n} a_{i,j}^k \quad (1)$$

with t being the count of the current trial, R being the reward derived from the ErrP-decoder class decision, with negative reward $R = -1$ for a classified error event and positive reward $R = +1$ for a classified non-error event, α being the learning rate, and $\sum_{k=1}^{n} a_{i,j}^k$ the occurrence count of action a_i in state s_j of the action sequence $k = (1, \ldots, n)$ executed by the robot in the current trial, with n depending on the subject's self-paced decision. Truncation and normalization was performed after adding the policy gradient $\alpha R \sum_{k=1}^{n} a_{i,j}^k$ to the parameters of the old policy π^t: parameter updates of π^{t+1} which exceeded the range $\{0, 1\} \in \mathbb{R}$ were truncated to 0 and 1, respectively, and all actions per state were then normalized to sum up to one. Based on prior experiments, the learning rate was empirically set to $\alpha = 0.1$ such that convergence could be reached relatively fast within a few policy updates. Fast convergence was preferred given the limited number of 50 trials per co-adaptation run[2]. The rationale behind including the empirical distribution of observed state-action pairs into the policy gradient was based on the assumption that more prominent state-action pairs are likely to contribute more to the subject's false or correct guess compared to less prominent state-action pairs. State-action pairs which occurred and hence were observed more often than others were as such more strongly reinforced (increase or decrease of corresponding state-action probability depending on R) compared to state-action pairs which occurred less prominently or never during the trial. This way, the policy is updated to promote correct guessing or in other words, to fit to the subject's belief by quantitatively taking into account the characteristic of the gazing behavior the subject has observed.

3. Results

Three out of 16 subjects did not meet the inclusion criterion defined in section 2.3.3: Offline decoder performance did not exceed the chance-level of 60.37% in either TPR, TNR or both in subject s05, s10, and s10 (supplementary table 3 (stacks.iop.org/JNE/15/066014/mmedia)). These subjects were excluded from subsequent data analyses. For the sake of full disclosure of the obtained data, individual results of excluded subjects are nevertheless reported and discussed separately.

3.1. ErrP decoding

Figure 3(a) shows the grand average ERP time-courses over channel Cz time-locked to the onset of LED-feedback presentation by the robot and their topographical distribution at specific time-points. The grand average difference (black line in figure 3(a)) showed the typical N2-P3-complex which has been reported consistently in the context of ErrPs (Ferrez and Millán 2008a, Chavarriaga *et al* 2014, Spüler and Niethammer 2015, Iturrate *et al* 2015). The negative deflection (N2-component, expected around 200–350 ms) was mostly pronounced frontocentrally around 300 ms post stimulus and the positive deflection (P3-component, expected around 250–500 ms) was mostly pronounced frontocentrally around 400 ms. The coefficient of determination based on channel Cz reached highest values of $r^2 = 0.09$ for 288 ms and $r^2 = 0.11$ for 394 ms averaged across all subjects ($n = 13$) which speaks in favor for a good overall separability of the data. Figure 3(b) shows a comparison of the grand average difference ERPs over Cz across calibration session (CALIB) and co-adaptation runs I, II, IV (CORL)[3] with high temporal resemblance between the experimental conditions.

The observations from the electrophysiological analysis were reflected in the single-trial classification performances (see figure 3(c) and supplementary table 3). The average offline ErrP decoder performance based on calibration data cross-validation was overall 80.2% ± 7.5% (ACC, overall accuracy), with TNR of 81.2% ± 7.7% and TPR of 79.2% ± 7.5%. ErrP online decoding performances were comparably high in accuracy (see figure 3(c) and supplementary table 3) with ACC = 84.2% ± 7.4% for CORL-I, ACC = 77.1% ± 12.1% for CORL-II, and ACC = 84.0% ± 10.6% for CORL-IV. The main difference observed in comparison to the offline cross-validation results was a higher decoding performance for non-error events (TNR: 86.5% ± 11.7%, 82.5% ± 18.3%, 90.4% ± 5.8% for CORL-I, II, IV, respectively) and a lower performance for error events (TPR: 75.3% ± 12.0%, 70.4% ± 15.4%, 74.4% ± 17.0% for CORL-I, II, IV, respectively). This performance bias was significant across subjects for all co-adaptation runs ($p = 0.026$, $p = 0.023$, $p = 0.002$, for CORL-I, II, IV, respectively; paired Wilcoxon signed rank test, $n = 13$). Online decoding accuracies were on average lower in CORL-II compared to CORL-I and CORL-IV. This was consistent across subjects as decoding accuracies differed significantly between CORL-I and CORL-II, and between CORL-II and CORL-IV; no statistically significant difference was found between CORL-I and CORL-IV ($p_{\mathrm{I-II}} = 0.031$, $p_{\mathrm{II-IV}} = 0.046$, $p_{\mathrm{I-IV}} = 0.600$; paired

[2] Analysis and results of policy convergence is reported in section 3.2.

[3] Please note that no results are reported for CO-RL-III, since no subject keypress responses (validation ground truth) were captured during this part of the experiment.

Table 2. Overview of individual results per subject in the order of columns from left to right: maximum coefficient of determination r^2 across all channels within period 150–550 ms for CALIB and CORL. Cross-validation ErrP-decoder accuracies based on CALIB data. Average online ErrP-decoder accuracies during co-adaptation runs CORL-I,-II, and -IV. Within-subject Pearson's spatiotemporal correlation coefficients between average difference ERP time courses of all channels (error minus non-error) of CALIB and CORL (average of all trials of I,II,IV) within period 150–550 ms.

	$r^2_{\max}$ CALIB	$\overline{r^2_{\max}}$ CORL	ACC CALIB (offline CV) (%)	$\overline{\text{ACC}}$ CORL (online acc.) (%)	corr2 (CALIB,CORL)
s03	0.20	0.31	69.3	86.7	0.67
s04	0.34	0.41	86.3	91.3	0.72
s06	0.36	0.20	82.0	72.7	0.74
s07	0.25	0.31	81.4	82.7	0.58
s08	0.43	0.48	85.7	87.3	0.87
s09	0.46	0.29	92.8	86.7	0.82
s11	0.17	0.09	68.9	65.3	0.29
s12	0.30	0.23	84.7	90.0	0.41
s14	0.39	0.27	88.7	77.3	0.72
s15	0.18	0.31	72.4	85.3	0.70
s16	0.17	0.31	73.3	74.7	0.49
s17	0.36	0.54	81.8	88.0	0.77
s18	0.39	0.24	78.9	74.7	0.53
AVG ± SD	**0.31 ± 0.10**	**0.31 ± 0.12**	**80.5 ± 7.5**	**81.8 ± 8.0**	**0.64 ± 0.17**
s05	0.12	0.12	50.5	50.7	−0.02
s10	0.10	0.19	64.4	28.7	0.13
s13	0.11	0.13	50.2	67.3	0.12

Wilcoxon signed rank test, $n = 13$). Possible explanations are discussed in section 4. The theoretical chance-level for online ErrP decoding per CORL is 62.0% (inverse cumulative binomial distribution with number of trials $nTrials = 50$; probability of success $p_{success} = 0.5$; confidence threshold $p = 0.05$) which was exceeded in all but three cases: s06/CORL-IV, s11/CORL-II, s14/CORL-II (see supplementary table 3).

Table 2 shows the overview of individual subject results. The separability of the ErrPs are expressed in form of the maximum coefficient of determination across all channels $r^2_{\max}$, separately for CALIB and the CORL data (all trials of CORL-I,-II, and -IV). The results show comparably low values of $r^2_{\max} \sim 0.11$ for the three subjects in which offline decoding performance did not exceed the chance-level threshold (s05, s10, s13), whereas all other subjects show higher $r^2_{\max}$ values. This indicates that calibration failed in these subjects, mainly due to their generally limited separability of ErrP responses (possible explanations are discussed in section 4). As expected, the overall ErrP-decoder offline cross-validation accuracies (ACC CALIB) and the online average decoding accuracies ($\overline{\text{ACC}}$ CORL) reflected the results obtained from the analysis of the coefficient of determination, with high separability resulting in higher decoding accuracies. Table 2 furthermore reports Pearson's spatiotemporal correlation coefficients between the average difference of ERP time courses of all channels (error minus non-error) within the period 150–550 ms (period in which the temporal features were extracted). The overall high correlation coefficients of average 0.64 ± 0.17 reflect high spatiotemporal resemblance and support the notion that the decoded ErrPs did not notably differ between CALIB and CORL experimental sessions, despite the different experimental conditions.

3.2. ErrP-based co-adaptation

To investigate the extent of co-adaptation between subject and robot, we analyzed the development of two behavioral measures in conjunction with the development of the policies during the co-adaptation runs: (1) Guessing performance—the development of the accuracy of correct guesses. This measure was expected to increase if both subject and robot converge to a consensus. (2) Gaze transitions until subject's decision—number of gaze transitions performed by the robot until the subject made a decision. This measure was expected to decrease as subjects and robot converge to a consensus. (3) Policy convergence—the policy change of trial-by-trial updates. This measure was expected to decrease if policies converge.

3.2.1. Efficacy: guessing performance. The subject has three objects to choose from and therefore chance-level was $p = \frac{1}{3}$. At the beginning of each co-adaptation run the robot's gaze policy was initialized with equal probabilities for all actions. This guaranteed a random guess in the first trial of all co-adaptation runs. Hence, if during the co-adaption runs, the subject's guessing performance exceeded chance-level, the robot's gaze policy must have been updated such that correct guessing was facilitated for the subject; vice-versa, if guessing performance did not increase above chance-level during the run, then updates in the robot's gaze policy did not facilitate the subject's task and/or were misleading. To investigate whether the guessing performance depended on the ErrP decoder performance during online operation, we computed Pearson's correlation coefficients between the overall guessing performance (percentage of correct guesses within one run) and the ErrP decoder accuracy (percentage of correctly classified trials) across all subjects for each co-adaptation run separately.

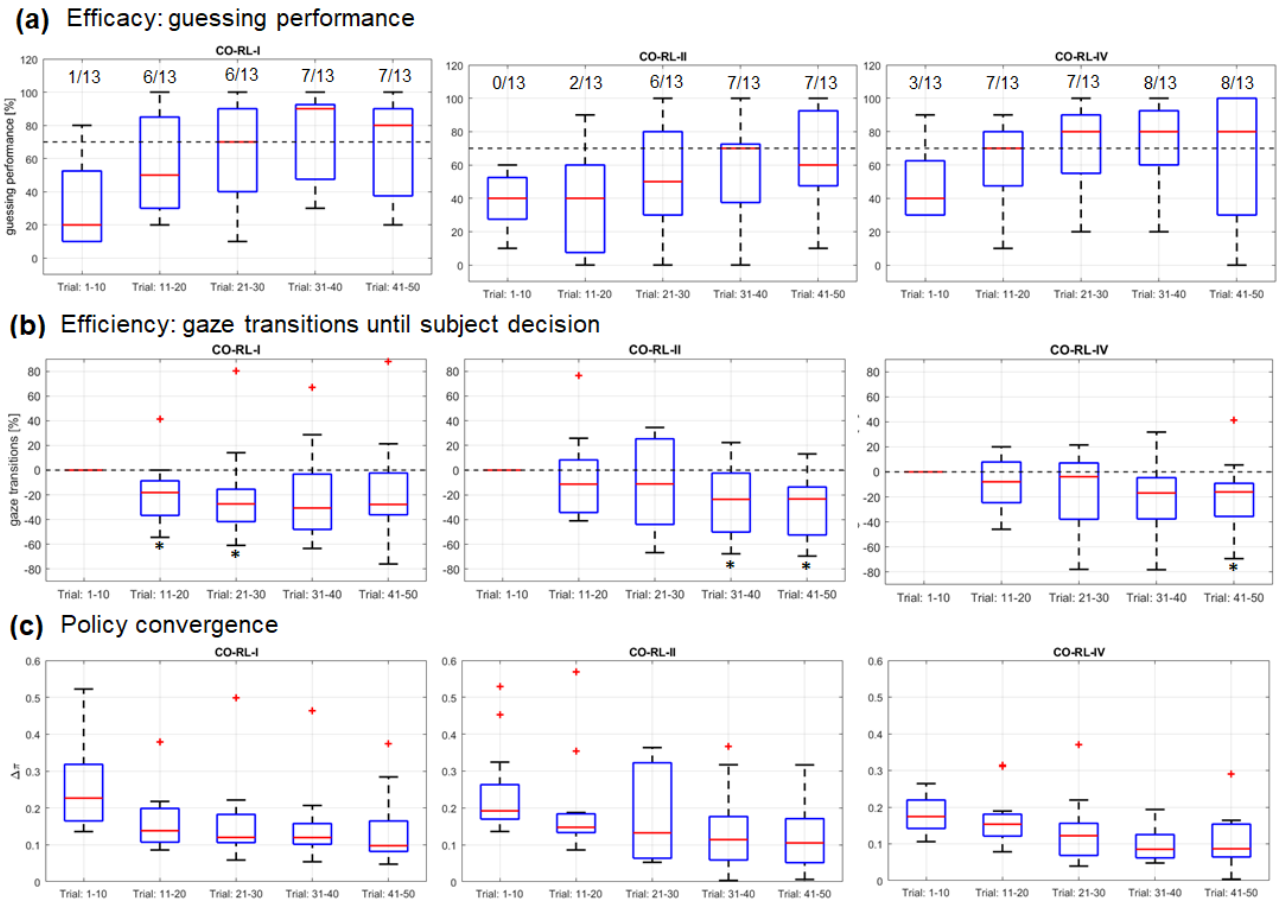

Figure 4. (a) Boxplot representation of guessing performance computed for consecutive segments of 10 trials, across subjects ($n = 13$). Different panels correspond to different co-adaptation runs. In all three runs, the median guessing performance increased to 70%–90%. The black dashed line indicates the threshold of the confidence interval of 70%. The numbers on top of the boxplots represent the ratio of subjects which exceeded the threshold of the confidence interval in the corresponding segment. (b) Boxplot representation of gaze transitions until subject decision computed within consecutive 10-trial segments, relative to the number of transition counted in the first segment (black dashed line). In all three co-adaptation runs, the median number of transitions decreased by 15%–27% with significant across subject deviations in some segments (boxes marked with a black asterisk). (c) Boxplot representation of policy change computed within consecutive 10-trial segments. In all three co-adaptation runs, the median policy change decreased from $\Delta\pi \sim 0.2$ in the first segment to $\Delta\pi \sim 0.1$ in the last segment, indicating policy convergence relative to increasing guessing performance.

Overall guessing performance correlated positively with the online decoding accuracies in all three co-adaptation runs with $r = .71$ ($p = 0.006$) for CORL-I, $r = .79$ ($p = 0.001$) for CORL-II, and $r = .47$ ($p = 0.1$) for CORL-IV (Pearson's correlation, $n = 13$). These results indicate that improvements in guessing performance depended on the ErrP decoder performance during online operation, e.g. high ErrP decoder performance fostering high guessing performance. As a result, those subjects in which the ErrP-decoder calibration performance resulted in below chance-level accuracies (s05, s10, s13), no notable improvements in guessing performance were observed in all co-adaptation runs of those subjects (compare supplementary tables 3 and 4). To investigate improvements of guessing performance over the course of co-adaptation runs, each run was partitioned into five segments of 10 trials each. Guessing performance was computed as percentage of correct guesses in each segment (the 5% confidence threshold is exceeded if $\geqslant$7 out of 10 trials were correct, one-sided binomial test with chance level $p = \frac{1}{3}$). Figure 4(a) shows across subject distributions of guessing performance from the start (trials: 1–10) until the end (trials: 41–50) of each co-adaptation run. The results show a median increase of guessing performance from initial chance-level up to 90% in CORL-I, 70% in CORL-II, and 80% in CORL-IV. In all three runs, the majority of subjects exceeded the threshold of the confidence interval (70%) at some point during the run. In CORL-I and CORL-II, in the fourth segment, and in CORL-IV already in the second segment. Despite the significant differences in ErrP-decoding performance (see section 3.1), no significant differences of overall guessing performance were observed between CORLs ($p_{\mathrm{I-II}} = 0.528$, $p_{\mathrm{II-IV}} = 0.250$, $p_{\mathrm{I-IV}} = 0.104$; paired Wilcoxon signed rank test, $n = 13$). Assuming a co-adaptation run to be 'successful' when guessing performance $\geqslant$70% in three subsequent segments (probability for exceeding by chance: $p = 7.6 * 10^{-6}$, one-sided binomial test with chance level $p = \frac{1}{3}$), then successful co-adaptation was achieved in 10 out of 13 subjects in at least one out of the three runs. Two subjects achieved 3/3 successful runs (s09, s12); two subjects achieved 2/3 successful runs (s03, s15), six subjects achieved 1/3 successful runs (s06, s07, s08, s14, s16, s18).

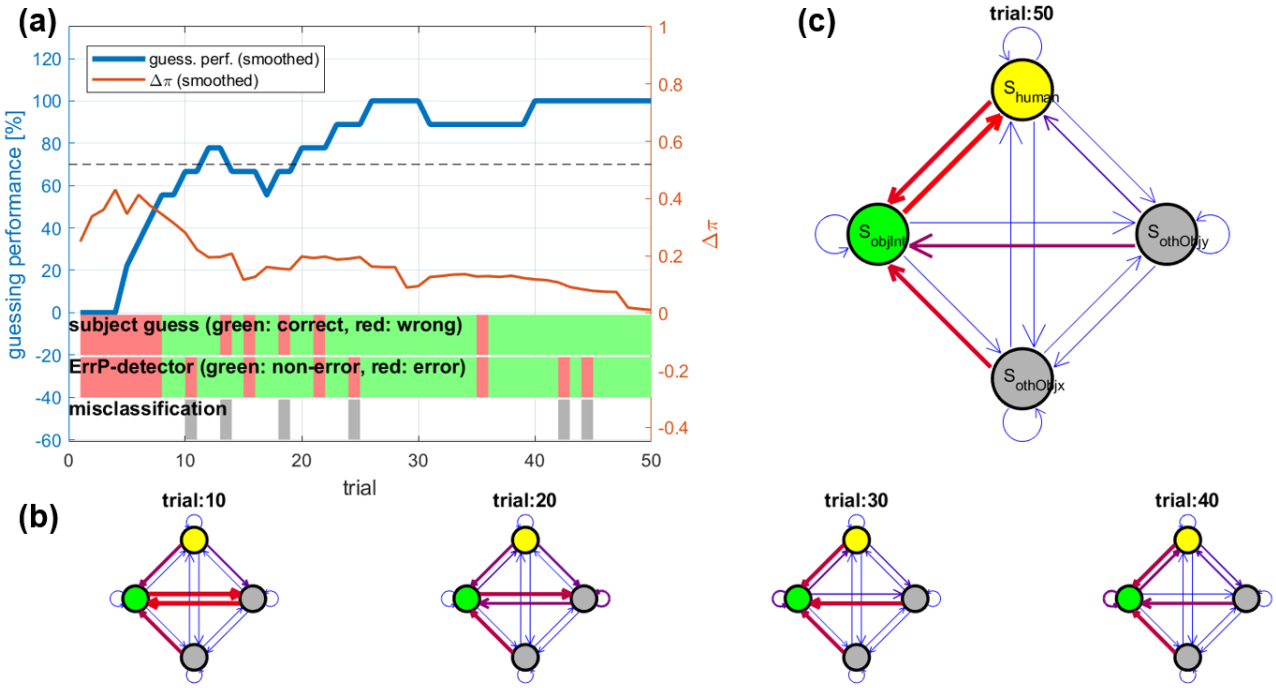

Figure 5. (a) The plots represent a smoothed representation (over 10 trials) of the development of guessing performance (blue line) together with the rate change of policy updates $\Delta\pi$ (orange line). The black dashed line represents the threshold of confidence (70%) for guessing performance. Single-trial subject guesses, corresponding ErrP-decoder classification decisions and misclassifications are illustrated below the plots. (b) Shows the gaze behavior policies after different numbers of iterations (after trial: 10, 20, 30, 40). (c) Shows the final policy at the end of the co-adaptation run. The states of the gazing policy are color-coded as follows: s_{human} (yellow), s_{ObjInt} (green), $s_{othObjx}$, $s_{othObjy}$ (grey). State-transitions with high probabilities are represented with thick red lines, low probabilities with thin blue lines. This particular example shows convergence towards the 'fixation' behavior around trial 30 and 'nodding' behavior towards the end of the co-adaptation run (see section 3.3). Individual results of all subjects are detailed in supplementary figures 1–64.

An exemplary successful co-adaptation run (s09/CORL-I) is visualized in figure 5. Individual results are detailed in supplementary table 4 and supplementary figures 1–64.

3.2.2. Efficiency: gaze transitions until subject decision. The absolute number of gaze transitions turned out to vary widely among subjects, even during the calibration session, ranging between 10 to 50 transitions (corresponding to a duration of robot action execution between ~4–20 s per trial) with average 13.0 ± 6.3 (CALIB), 15.7 ± 7.0 (CORL-I), 15.0 ± 6.9 (CORL-II), and 14.8 ± 10.7 (CORL-IV). Therefore, the number of gaze transitions until subject decision was analyzed by partitioning each co-adaptation run into five segments of 10 trials each (in accordance with the analysis of guessing performance) and counting the number of gaze transitions within each of these 10-trial segments relative to the number of transitions occurring during the first 10-trial segment. The results are depicted in figure 4(b): in all three runs, the median number of gaze transitions decreased by 15%–27% relative to the first segment with across subject significant deviation in some segments ($p < 0.05$, one-sample Wilcoxon signed rank test). In CORL-I the number of gaze transitions decreased by 27.6% (median of percent reduction calculated across subjects); in CORL-II by 19.2% and in CORL-IV by 15.6%. This result illustrates that during the co-adaptation runs, subjects not only became more precise in guessing, but also on average faster in deciding about the robot's selected object. This suggests that the robot's gaze behavior adapted in a way that was generally easier and quicker understood by participants. The absolute number of gaze transitions until subject decision is given in supplementary table 5.

3.2.3. Policy convergence. To quantify policy convergence, the difference between subsequent policy iterations was computed for each subject and CORL individually. This was carried out by determining the value of the state-action pair with the maximum difference between subsequent policy iterations, termed as the policy change after trial k: $\Delta\pi^k = \max(|\pi^{k-1} - \pi^k|)$. In accordance with the previous analyses of behavioral measures, policy convergence was analyzed by partitioning each co-adaptation run into five segments of 10 trials each and averaging $\Delta\pi$ within each of these 10-trial segments. The results are depicted in figure 4(c): in all three runs, the median of $\Delta\pi$ across subjects decreased steadily from the first until the last segment from an initial median of $\Delta\pi \sim 0.2$ to a final median of $\Delta\pi \sim 0.1$. This indicates that policies were on average converging relative to an increasing guessing performance. The results further indicate that not all policies converged within 50 trials as the median policy change was still $\Delta\pi \sim 0.1$ in the last segment of all three runs. These findings are further discussed in section 4. Policy convergence for an exemplary successful

co-adaptation run (s09/CORL-I) is visualized in figure 5. Individual results of policy convergence are detailed in supplementary figures 1–64.

3.3. Emergence of gaze behavior

In addition to assessing the development of co-adaptation, we were also interested in the nature of the gaze behavior which emerged during the co-adaptation runs. This analysis allowed for a qualitative assessment of CORL-III in which subjects were not requested to explicitly indicate their guesses via keypress responses in comparison with the other co-adaptation runs. Since participants were not instructed to follow a particular strategy/policy, any gaze behavior was considered acceptable and denoted as useful if it helped the subject to perform better and faster in guessing the robot's selected object. Figure 6(a) shows an overview of the learned policies for the 13 subjects and all co-adaptation runs, including CORL-III. Successful co-adaption was expected to be reflected in policy convergence towards the end of the run. Therefore the average of the policies of the last 10 trials (trial: 41–50) were depicted, with thick red lines representing high probabilities and thin blue lines low probabilities. The policies which emerged from successful co-adaptation runs are highlighted with a blue frame; the guessing performances of the corresponding last 10 trials are furthermore depicted next to the average policy. By qualitative visual inspection, we identified two different recurring policies which are furthermore termed 'fixation' and 'nodding' behavior (see figure 6(b)). The 'fixation' policy led to gaze behavior in which the robot tended to fixate the selected object. Example cases are s03/CORL-I, s07/CORL-II, s15/CORL-I. In the 'nodding' policy the robot was gazing alternatingly between the subject and the selected object in a nodding-type fashion. Examples are s09/CORL-I, s12/CORL-II, s18/CORL-IV. Also in CORL-III, the robot's gaze behavior converged in a few cases to one of the two identified policies, e.g. 'fixation' behavior in s08 and s18, and 'nodding' behavior in s16 (figure 6(a)). These cases indicate that participants explicitly indicating their decision (as in CORL-I, -II, -IV) was not required for successful co-adaptation, and suggests that the ErrP-based method presented here also worked based on covert beliefs/decisions without explicit actions linked to the decisions. Convergence to the 'fixation' behavior was expected, since it is very similar to the pre-programmed policy used during the calibration task; subjects likely could have used it as a proxy. The 'nodding' behavior was however unexpected, since it had not occurred before during the calibration session and subjects could therefore not use it as a proxy. Interestingly, the number of cases associated with convergence to the 'nodding' policy (7) were approximately on par with those with convergence to the 'fixation' policy (8). The 'nodding' behavior may have emerged from subjects gradually finding it useful and in result having adapted to and positively reinforced it. This observation retrospectively confirms co-adaptation between subject and robot, since if subjects were instructed to teach the robot a specific behavior, previously unexpected behavior is unlikely to emerge.

4. Discussion

ErrPs, decoded from human subjects' brain activity in real-time during HRI, might be useful in the future to adapt the behavior of artificial agents, such as robots, to better align with human expectations, needs and conventions. We understand our study as a logical extension of previous works (Iturrate *et al* 2015, Salazar-Gomez *et al* 2017, Kim *et al* 2017) which demonstrated the potential of using ErrPs as a teaching signal for robot skill learning. In contrast, our experimental paradigm featured a scenario in which there was no explicit 'optimal' or 'correct' behavior the robot was required to adapt to, but where mutual adaptation between human and robot was permitted; the 'optimal' robot policy had to be negotiated between both parties in a co-adaptive fashion. This introduced a considerable level of uncertainty and complexity into the experimental setup as subjects could not follow a specific task or proxy. With this relaxation of constraints in the experimental setup, we aimed at validating the usability of ErrPs as an implicit feedback signal to improve HRI where adaptation is possible from both interaction partners. Despite the uncertainty and complexity introduced, we observed significant improvements in interaction performance across participants over the course of individual co-adaptation runs, as indicated by behavioral measures of efficacy and efficiency: The average percentage of correct guesses (efficacy) increased from the initial chance-level (~33%) to 70%–90% within 10–40 trials (corresponding to 1–4 min), median across subjects. Additionaly, the number of gaze transitions made by the robot before the participant indicating his/her guess (efficiency), relative to the corresponding number in the beginning of the co-adaptation run, decreased on average by 15%–27%. Hence, adaption of robot's policy, based on the ErrPs collected from the human interaction partner, was accompanied by a higher performant and more efficient interaction.

Online single-trial ErrP decoding performance was on average 81.8% $\pm$ 8.0% across 13 subjects which is comparable to previously reported ErrP classification performances used for closed-loop adaptation of robotic systems: Iturrate *et al* (2015) obtained online decoding accuracies around 74% across 12 subjects using temporal features combined with LDA classification. Salazar-Gomez *et al* (2017) obtained online decoding accuracies around 65% across four subjects using correlation-based features and covariance features based on spatially filtered EEG signals using xDAWN (Rivet *et al* 2009). Based on post-hoc offline analyses, they reported however about the presence of a secondary ErrP for which they estimated a theoretical online decoding performance around 80%. The most recent work by Kim *et al* (2017) reported high online decoding performances of balanced accuracy around 90% across seven subjects using temporal features after xDAWN spatial filtering and classification using linear support vector machines (SVM). They explain their high performance being mainly a result of their data augmentation approach based on decoding ErrPs in two separate time windows (instead of just one). A limitation observed from the online single-trial classification results of the present study (section 3.1) is a consistent and significant across subjects

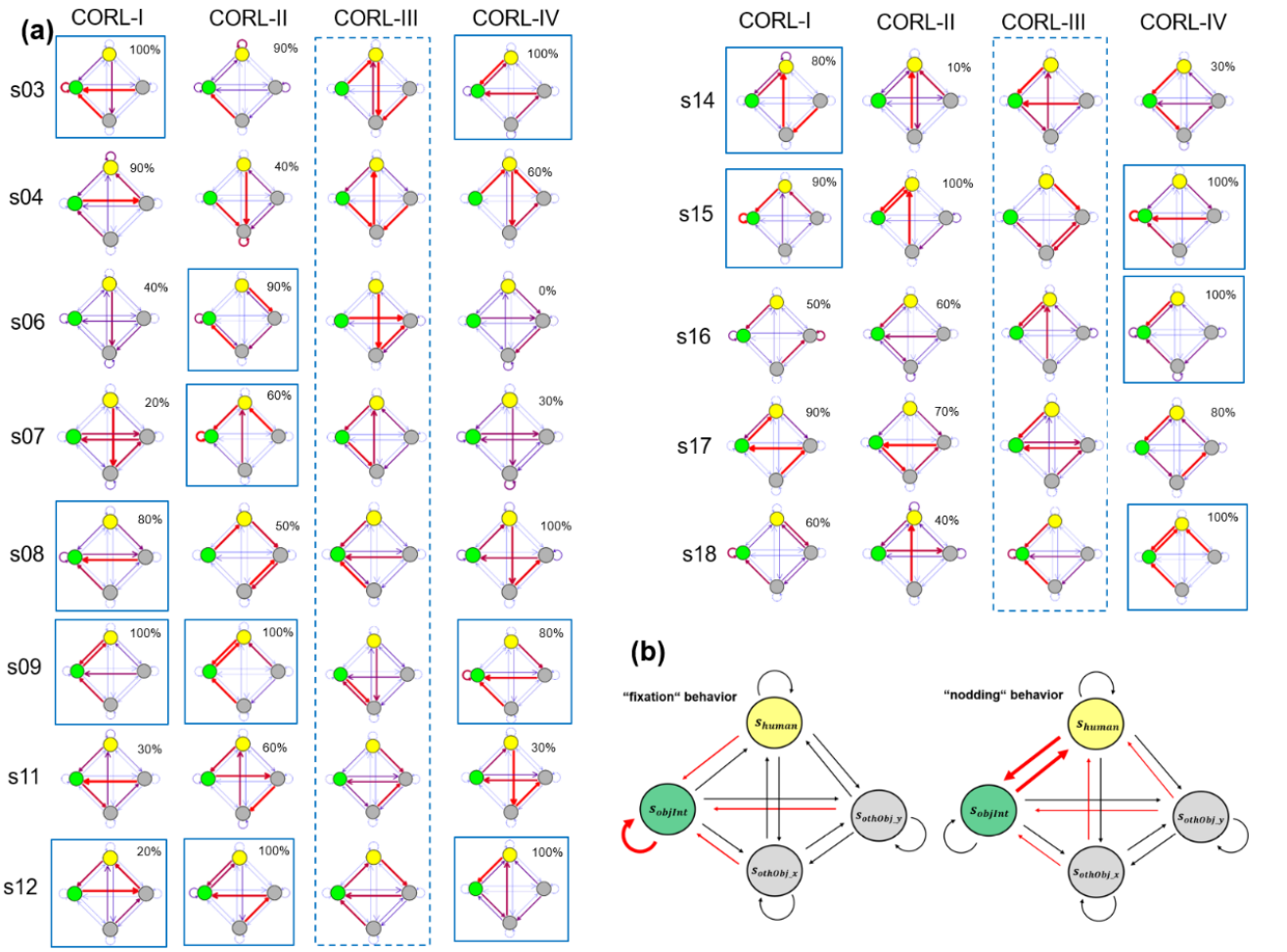

Figure 6. (a) Overview of gaze policies averaged across the last 10 trials for 13 subjects and all co-adaptation runs. Successful co-adaptation runs are highlighted with blue frames. The guessing performance during the corresponding last 10 trials is depicted next to the policies. (b) Identified recurring policies: 'fixation' behavior in which the robot tended to fixate the selected object (examples: s03/CORL-I, s15/CORL-I, s08/CORL-III, s18/CORL-III); 'nodding' behavior in which the robot alternatingly gazed at the subject and the selected object in a nodding-type fashion (examples: s09/CORL-II, s16/CORL-III, s18/CORL-IV).

bias towards the non-error class. This classification bias has been reported consistently in the context of ErrP decoding and related to the typical design of calibration protocols with unbalanced number of samples per class (Chavarriaga *et al* 2014). As class-balancing was performed in the present study, the systematic bias was likely related to the limited number of 150 samples used for ErrP-decoder calibration. Although the use of a regularized LDA may have partially counteracted this (Schäfer and Strimmer 2005), the use of *a priori* information from other subjects (Iturrate *et al* 2011) or upsampling the minority class, instead of downsampling the majority class may have helped improving online decoding and are recommended modifications for future works. ErrP decoder calibration resulted in chance-level performance in three subjects (s05, s10, s13). The results reported in table 2 are informative in that they show low $r^2_{\max}$ values compared to the remaining subjects, indicating generally limited separability of their ErrP responses. A post-hoc visual inspection of the raw EEG data showed comparably strong artifact contamination in s10 (mainly noisy channels) and in s13 (mainly slow DC drifts); the data of s05 on the other hand was largely unaffected by artifacts. This suggests that the low calibration performance was mainly due to the technical setup and could have been resolved by repeating the experiment on a different day or by adding automatic artifact rejection to the modeling procedure. Why calibration failed in s05 is currently unexplained; one possibility could be that this subject had been insufficiently concentrated on or engaged in the task. Across subjects, ErrP online decoding performance was significantly lower in CORL-II than in CORL-I and CORL-IV. Notably, this systematic performance drop was temporary and fully recovered toward the end of the experiment in CORL-IV. Therefore it is hypothesized that the observed performance drop is barely related to technical reasons, but rather to the subject's concentration/task engagement level. CORL-II was a direct repetition of CORL-I, which might have had a negative effect on the subject's motivation and engagement. On the other hand, CORL-IV followed CORL-III; the intermediate variation of the experimental protocol with CORL-III might have had a positive effect on the subject's engagement during CORL-IV. Despite noticeable differences in the median guessing performance (figure 4(a)), the systematic drop in ErrP online decoding performance had no significant effect on the overall guessing performance in CORL-II.

The ErrP-decoder performance played an integral role in successful co-adaptation, as indicated by positive correlations between overall guessing performance and ErrP decoder performance during online operation. Furthermore, on average, policies converged in relation to increasing guessing performance (see figures 4(a) and (c)) as indicated by a median decrease of the policy rate change over the course of co-adaptation runs. This supports the functionality of the policy adaptation approach here adopted. We observed, however, a number of cases with failed co-adaptation despite high ErrP decoder performance (ACC $> 75\%$). These cases might be due to unknown human-related factors, e.g. variations in engagement in the task, attention variations, or variations in interpretation of the experiment. From a technical perspective, the policy adaptation approach used may have influenced the stability of co-adaptation as well. In some of these failed-cases there were temporary increases in guessing performance followed by decreases (unlearning), indicating temporary, but unstable learning (exemplary cases are s14/CORL-IV, s17/CORL-II, see supplementary figures 36 and 46). The learning approach here adopted does not enforce convergence to a global optimum. This has the advantage that while converging to one policy, bifurcations to other policies remain possible. This flexibility might be particularly important in the context of co-adaptation as changes of the human strategy are likely and imaginable in the sense that a policy which was previously optimal to the subject is neglected and replaced by a different optimal policy. Exemplary cases supposedly showing such policy re-adaptations are s08/CORL-IV, s09/CORL-II, and s14/CORL-I (supplementary figures 20, 22, and 33). These cases show initial convergence interrupted by periods of increased policy changes and subsequent secondary convergence. On the other hand, this flexibility, in combination with a learning rate parametrized to promote quick learning, may have encouraged instabilities or quick unlearning, as the outcome of single trials interfered with the learning process (e.g. sensibility to ErrP-decoder misclassifications). One possible way of stabilizing the policy adaptions would be to use an adaptive learning rate based on the past rewards (ErrP-decoder decisions), e.g. decreasing the learning rate in case of increasing number of past non-error events. However, whether, and to what extent a systematic control of the learning process is recommendable in the context of ErrP-based human-agent co-adaptation remains an open question for future investigations. In our experiment the learning process was limited to 50 iterations (trials), which turned out insufficient for drawing definite conclusions about the co-adaptation process in the long run. Therefore, an entry point for follow-up studies is most importantly the investigation of the dynamic effects of co-adaptation for longer periods or during continuing interaction.

The analysis of emergence of gaze policies revealed that in most successful co-adaptation runs the robot's gaze behavior converged to either one of two different policies: 'fixation' and 'nodding' behavior. While the 'fixation' behavior was expected as it closely resembles the gaze behavior during the calibration session, the 'nodding' behavior, in contrast, was not expected. Although both behaviors are noticeably different and may be interpreted as conveying different meanings, an alternative interpretation is that both are consistent in that the target object is attended to more often than others. In that sense, it is likely that the type of behavior to which the system converged to depended on whether the alternating transitions between s_{human} and s_{objInt} or the transition $s_{objInt \rightarrow objInt}$ were sampled more often in an early stage of the co-adaptation run. Either way of interpreting the emergence of the two types of behaviors supports the hypothesis of co-adaptation, since for the subjects both strategies seemed valid despite having had no explicit exposure to the 'nodding' behavior before the start of the co-adaptation runs. On that note, one may argue about why just two different behaviors emerged from the interaction, given that the manifold of imaginable and possibly valid strategies is much bigger (e.g. a slightly more complicated gaze pattern or a consistent logical swap of the target object with one of the other objects). This observation may be related to constraints in human information processing and learning of more complex statistical patterns but remains an open question for future investigations. Further exploration of how robot behavioral policies, as in this case the robot's gazing policy, develop during such interaction will provide useful insights for improving the technical implementation of ErrP-based mediation of human-robot co-adaptation and may likewise provide insights about human information processing and learning.

Our experiment featured a rather synthetic and highly structured HRI scenario. This was necessary for an initial proof of concept of our approach. Most simplifications and procedural constraints were introduced to ensure reliable decoding of ErrPs from EEG signals as well as for the purpose of clean validation: For instance, our experiment was designed in a way as if no explicit human feedback (key-press response) was available, to allow quantitative validation of ErrP-decoder performance as well as improvements in interaction performance indicative of co-adaptation. In this regard, the additional human feedback in form of key-press responses served only as a ground truth measure for post-hoc validation. Comparison of learned gaze policies between CORL-III (no explicit feedback) and the other conditions CORL-I,-II,-IV (with explicit feedback) suggested however that the explicit human feedback was not a pre-requisite for successful co-adaptation. However, despite the few examples of successful co-adaptation in CORL-III, no definite conclusions can be drawn from this part of the study. Notably, the results reported in supplementary table 5 suggest that 15 gaze transition may have not been enough for all subjects to build up an estimate about the target object with a level of confidence high enough to elicit observable (and decodeable) ErrPs. The design choice of 15 gaze transitions may have hampered the outcome of CORL-III; a follow-up study with a design focused on the rationale and motivation of CORL-III is required to consolidate the preliminary findings of this part of the study. Furthermore, we used a perceptually simple symbolic feedback in form of a flashing LED for communicating the robot's selected object to the subject. This was necessary as earlier works demonstrated limited ErrP decodeability in response to perceptually more complex or gradually unfolding stimuli

(Omedes *et al* 2015, Ehrlich and Cheng 2016, Welke *et al* 2017, Dias *et al* 2018). Subject-dependent individual ErrP-decoders had to be calibrated. This is a typical procedure in the field of EEG-based BCI research (Wolpaw *et al* 2002) and necessary because of typically high inter-subject variations of EEG signals and responses (Lotte and Guan 2010). Even though ErrPs have been found to largely resemble in terms of spatiotemporal activity patterns, earlier works have highlighted task-dependent ErrP signal variations that can negatively affect decoding performance when applying decoders across task (Iturrate *et al* 2013). Therefore, we employed a calibration protocol which contextually resembled the co-adaptation runs. Although ErrPs have been widely recognized as a useful response to harvest from human subjects for improving human-machine interaction, the design constraints and simplifications introduced in the present study illustrate the current challenge of the method's straightforward applicability. Research efforts on different ends, such as on improving decoding performance (Omedes *et al* 2015, Kim *et al* 2017), on practicality of the EEG setups (Ehrlich *et al* 2017) and decoder calibration (Iturrate *et al* 2011, Kim and Kirchner 2016), and on the observability of ErrPs in response to different stimuli and varying scenarios (Ehrlich and Cheng 2016, Welke *et al* 2017, Behncke *et al* 2018, Omedes *et al* 2018) are required to push ErrP-decoding towards more widespread applicability.

5. Conclusion

In this paper, we experimentally demonstrated the usability of EEG-based ErrPs as a feedback signal for mediating co-adaptation in HRI. Our study featured a simplified HRI scenario in which successful interaction depended on co-adaptive convergence to a consensus between subject and robot. ErrPs were decoded online from subjects' ongoing EEG signals with an avg. accuracy of $81.8 \pm 8.0\%$ and utilized for adaptations of the robot behavior, while the subject adapted to the robot by reflecting upon its behavior. Adaptation of the robot behavior was realized with an episode update strategy using ErrPs as a delayed reward feedback signal for the past sequence of robot actions. Successful co-adaptation was demonstrated by significant improvements in interaction efficacy and efficiency across subjects and by the robot behavioral policies that emerged during the interaction.

Acknowledgments

We thank Ana Alves-Pinto, Sae Franklin, and Pablo Lanillos for helpful comments on experimental design and data analysis. We thank the anonymous reviewers for their detailed comments and references, which have led to significant clarification of the work in this paper. This research was partially supported by Deutsche Forschungsgemeinschaft (DFG) through the International Graduate School of Science and Engineering (IGSSE) at the Technical University of Munich (TUM).

Competing interests

The authors declare no conflict of interest, nor competing financial interests.

Ethics approval and consent taking

This work was approved by the ethics commission of the Faculty of Medicine, Technical University of Munich under the reference number 236/15s. Consent to participate and publish was obtained from the participants in verbal and written form.

Availability of data and material

Data and material was not made publicly available but can be obtained from the corresponding author.

ORCID iDs

Stefan K Ehrlich https://orcid.org/0000-0002-3634-6973
Gordon Cheng https://orcid.org/0000-0003-0770-8717

References

Alexander W H and Brown J W 2011 Medial prefrontal cortex as an action-outcome predictor *Nat. Neurosci.* **14** 1338

Behncke J, Schirrmeister R T, Burgard W and Ball T 2018 The role of robot design in decoding error-related information from EEG signals of a human observer (arXiv:1807.01597)

Blankertz B, Dornhege G, Schafer C, Krepki R, Kohlmorgen J, Muller K R, Kunzmann V, Losch F and Curio G 2003 Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 127–31

Blankertz B, Lemm S, Treder M, Haufe S and Müller K R 2011 Single-trial analysis and classification of ERP components—a tutorial *NeuroImage* **56** 814–25

Botvinick M M, Braver T S, Barch D M, Carter C S and Cohen J D 2001 Conflict monitoring and cognitive control *Psychol. Rev.* **108** 624

Chavarriaga R and Millán J D R 2010 Learning from EEG error-related potentials in noninvasive brain–computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **18** 381–8

Chavarriaga R, Sobolewski A and Millán J D R 2014 Errare machinale est: the use of error-related potentials in brain-machine interfaces *Frontiers Neurosci.* **8** 208

Dias C L, Sburlea A I and Müller-Putz G R 2018 Masked and unmasked error-related potentials during continuous control and feedback *J. Neural Eng.* **15** 036031

Ehrlich S and Cheng G 2016 A neuro-based method for detecting context-dependent erroneous robot action *2016 IEEE-RAS 16th Int. Conf. on Humanoid Robots (Humanoids)* (IEEE) pp 477–82

Ehrlich S, Alves-Pinto A, Lampe R and Cheng G 2017 A simple and practical sensorimotor EEG device for recording in patients with special needs *Neurotechnix2017, CogNeuroEng 2017, Symp. on Cognitive Neural Engineering* pp 73–9

Falkenstein M, Hoormann J, Christ S and Hohnsbein J 2000 ERP components on reaction errors and their functional significance: a tutorial *Biol. Psychol.* **51** 87–107

Ferrez P W and Millán J D R 2005 You are wrong!—automatic detection of interaction errors from brain waves *Proc. 19th Int. Joint Conf. on Artificial Intelligence (No. EPFL-CONF-83269)*
Ferrez P W and Millán J D R 2008a Error-related EEG potentials generated during simulated brain–computer interaction *IEEE Trans. Biomed. Eng.* **55** 923–9
Ferrez P W and Millán J D R 2008b Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy *Proc. 4th Int. Brain–Computer Interface Workshop and Training Course (No. CNBI-CONF-2008-004)* pp 197–202
Friedman J H 1989 Regularized discriminant analysis *J. Am. Stat. Assoc.* **84** 165–75
Garrido M I, Kilner J M, Stephan K E and Friston K J 2009 The mismatch negativity: a review of underlying mechanisms *Clin. Neurophysiol.* **120** 453–63
Gouaillier D, Hugel V, Blazevic P, Kilner C, Monceaux J, Lafourcade P, Marnier B, Serre J and Maisonnier B 2008 The nao humanoid: a combination of performance and affordability *CoRR* (arXiv:0807.3223)
Gullapalli V, Franklin J A and Benbrahim H 1994 Acquiring robot skills via reinforcement learning *IEEE Control Syst.* **14** 13–24
Hadar U, Steiner T J, Grant E C and Rose F C 1984 The timing of shifts of head postures during conversation *Hum. Mov. Sci.* **3** 237–45
Holroyd C B and Coles M G 2002 The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity *Psychol. Rev.* **109** 679
Homan R W, Herman J and Purdy P 1987 Cerebral location of international 10–20 system electrode placement *Electroencephalogr. Clin. Neurophysiol.* **66** 376–82
Iturrate I, Montesano L and Minguez J 2010 Single trial recognition of error-related potentials during observation of robot operation *2010 Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 4181–4
Iturrate I, Montesano L, Chavarriaga R, Millán J D R and Minguez J 2011 Minimizing calibration time using inter-subject information of single-trial recognition of error potentials in brain–computer interfaces *Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (Boston, MA, 2011)* pp 6369–72
Iturrate I, Montesano L and Minguez J 2013 Task-dependent signal variations in eeg error-related potentials for brain–computer interfaces *J. Neural Eng.* **10** 026024
Iturrate I, Chavarriaga R, Montesano L, Minguez J and Millán J D R 2015 Teaching brain–machine interfaces as an alternative paradigm to neuroprosthetics control *Sci. Rep.* **5** 13893
Kim S K and Kirchner E A 2016 Handling few training data: classifier transfer between different types of error-related potentials *IEEE Trans. Neural Syst. Rehabil. Eng.* **24** 320–32
Kim S K, Kirchner E A, Stefes A and Kirchner F 2017 Intrinsic interactive reinforcement learning–Using error-related potentials for real world human-robot interaction *Sci. Rep.* **7** 17562
Kothe C 2014 Lab streaming layer (LSL) (https://github.com/sccn/labstreaminglayer) (Accessed: 26 October 2015)
Kreilinger A, Neuper C and Müller-Putz G R 2012 Error potential detection during continuous movement of an artificial arm controlled by brain–computer interface *Med. Biol. Eng. Comput.* **50** 223–30
Lanillos P, Ferreira J F and Dias J 2015 Designing an artificial attention system for social robots *2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)* pp 4171–8
Lotte F and Guan C 2010 Learning from other subjects helps reducing brain–computer interface calibration time *2010 IEEE Int. Conf. on Acoustics Speech and Signal Processing (ICASSP)* (IEEE) pp 614–7
Miltner W H, Braun C H and Coles M G 1997 Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a 'generic' neural system for error detection *J. Cogn. Neurosci.* **9** 788–98
Oliveira F T, McDonald J J and Goodman D 2007 Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations *J. Cogn. Neurosci.* **19** 1994–2004
Omedes J, Iturrate I, Minguez J and Montesano L 2015 Analysis and asynchronous detection of gradually unfolding errors during monitoring tasks *J. Neural Eng.* **12** 056001
Omedes J, Schwarz A, Müller-Putz G R and Montesano L 2018 Factors that affect error potentials during a grasping task: toward a hybrid natural movement decoding BCI *J. Neural Eng.* **15** 046023
Peirce J W 2007 PsychoPy—psychophysics software in Python *J. Neurosci. Methods* **162** 8–13
Peters J and Schaal S 2008 Reinforcement learning of motor skills with policy gradients *Neural Netw.* **21** 682–97
Renard Y, Lotte F, Gibert G, Congedo M, Maby E, Delannoy V, Bertrand O and Lécuyer A 2010 Openvibe: an open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments *Presence* **19** 35–53
Ridderinkhof K R, Ullsperger M, Crone E A and Nieuwenhuis S 2004 The role of the medial frontal cortex in cognitive control *Science* **306** 443–7
Salazar-Gomez A F, DelPreto J, Gil S, Guenther F H and Rus D 2017 Correcting robot mistakes in real time using eeg signals *2017 IEEE Int. Conf. on Robotics and Automation (ICRA)* (IEEE) pp 6570–7
Sallet J, Quilodran R, Rothé M, Vezoli J, Joseph J P and Procyk E 2007 Expectations, gains, and losses in the anterior cingulate cortex *Cogn. Affect. Behav. Neurosci.* **7** 327–36
Schäfer J and Strimmer K 2005 A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics *Stat. Appl. Genet. Mol. Biol.* **4**
Schalk G, Wolpaw J R, McFarland D J and Pfurtscheller G 2000 EEG-based communication: presence of an error potential *Clin. Neurophysiol.* **111** 2138–44
Schlögl A, Keinrath C, Zimmermann D, Scherer R, Leeb R and Pfurtscheller G 2007 A fully automated correction method of EOG artifacts in EEG recordings *Clin. Neurophysiol.* **118** 98–104
Schmidt N M, Blankertz B and Treder M S 2012 Online detection of error-related potentials boosts the performance of mental typewriters *BMC Neurosci.* **13** 19
Sidner C L, Kidd C D, Lee C and Lesh N 2004 Where to look: a study of human-robot engagement *Proc. 9th Int. Conf. on Intelligent User Interfaces* (ACM) pp 78–84
Spüler M, Rosenstiel W and Bogdan M 2012a Online adaptation of a c-VEP brain–computer interface (BCI) based on error-related potentials and unsupervised learning *PLoS One* **7** e51077
Spüler M, Bensch M, Kleih S, Rosenstiel W, Bogdan M and Kübler A 2012b Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI *Clin. Neurophysiol.* **123** 1328–37
Spüler M and Niethammer C 2015 Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity *Frontiers Hum. Neurosci.* **9** 155
Sutton R S and Barto A G 1998 *Reinforcement Learning: an Introduction* vol 1 (Cambridge, MA: MIT Press)
Rivet B, Souloumiac A, Attina V and Gibert G 2009 xDAWN algorithm to enhance evoked potentials: application to brain–computer interface *IEEE Trans. Biomed. Eng.* **56** 2035–43
Ullsperger M, Danielmeier C and Jocham G 2014 Neurophysiology of performance monitoring and adaptive behavior *Physiol. Rev.* **94** 35–79
van Schie H T, Mars R B, Coles M G and Bekkering H 2004 Modulation of activity in medial frontal and motor cortices during error observation *Nat. Neurosci.* **7** 549
Welke D, Behncke J, Hader M, Schirrmeister R T, Schönau A, Eßmann B, Müller O, Burgard W and Ball T 2017 Brain responses during robot-error observation (arXiv:1708.01465)
Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91