International Journal of Social Robotics (2024) 16:185–196
https://doi.org/10.1007/s12369-023-01073-2

Type of Education Affects Individuals' Adoption of Intentional Stance Towards Robots: An EEG Study

Cecilia Roselli[1] · Uma Prashant Navare[1,2] · Francesca Ciardo[1] · Agnieszka Wykowska[1]

Accepted: 20 October 2023 / Published online: 10 November 2023

Abstract

Research has shown that, under certain circumstances, people can adopt the *Intentional Stance* towards robots and thus treat them as intentional agents. Previous evidence showed that there are factors at play in modulating the Intentional Stance, for example individuals' years of education. In the present study, we aimed at investigating whether, given the same years of education, participants' type of formal education- in terms of theoretical background- affected their adoption of the Intentional Stance.

To do so, we recruited two samples of participants varying in their type of formal education, namely, a sample of participants comprised individuals with a background in robotics, whereas the other comprised individuals with a background in psychotherapy. To measure their likelihood of adopting the Intentional Stance, we asked them to complete the InStance Test (IST). To do it at the neural level, we recorded their neural activity during a resting state via electroencephalography (EEG).

Results showed that therapists attributed higher IST scores of intentionality to the robot than roboticists, i.e., they were more likely to attribute Intentional Stance to explain robot's behaviour.

This result was mirrored by participants' EEG neural activity during resting state, as we found higher power in the gamma frequency range (associated with mentalizing and the adoption of Intentional Stance) for therapists compared to roboticists.

Therefore, we conclude that the type of education that promotes mentalizing skills increases the likelihood of attributing intentionality to robots.

Keywords Intentional Stance · Education · Resting State

1 Introduction

1.1 Mentalizing and Intentional Stance

Mentalizing is a fundamental cognitive capability of humans: the ability to mentalize allows for successful navigation through complex social life [1]. The term "mentalizing" usually refers to humans' process of reasoning about their own and others' mental states, such as action goals and intentions, as well as higher-level states such as feelings, attitudes, and beliefs [1]. In this paper, we define mentalizing as a process of reasoning about a *particular* mental state driving a *particular* behavior, in a *specific* context, as in, for example: "*she called her mother on the phone, because she wanted to ask for advice*". As this example shows, mentalizing allows for understanding and predicting behavior of others.

However, in order to apply the mechanism of mentalizing, one needs to first adopt the Intentional Stance [2–5] towards the agent, whose behavior is being explained/predicted. That is, one needs to assume that the agent has the *capacity* for mental states (i.e., is an intentional agent with beliefs, desires and intentions). This is obvious and default

Cecilia Roselli and Uma Prashant Navare equally contributed to the work.

✉ Agnieszka Wykowska
Agnieszka.Wykowska@iit.it

[1] Social Cognition in Human-Robot Interaction, Italian Institute of Technology, via Enrico Melen 83, Genoa 16152, Italy

[2] Faculty of Science and Engineering, Manchester University, Manchester M15GF, UK

in the case of humans, but not necessarily in case of artificial agents, such as robots.

Thus, the concepts of *mentalizing* and *Intentional Stance* are similar, but not equivalent. While the Intentional Stance is a more general strategy or “attitude” adopted towards others when trying to explain/predict their behaviour, mentalizing is the active process of reasoning about a given observed behaviour, its causes and consequences. A clear example for this distinction comes from how mentalizing has been operationalized in laboratory settings, namely the false beliefs task of Wimmer and Perner [6]. It is a task typically used in developmental psychology to assess individuals’ mentalizing abilities. This task requires the capacity to understand that what other person “knows” is not necessarily what oneself knows, thus it requires taking perspective of others. In this task, participants typically observe a protagonist putting an object in a location (e.g., a basket). Participants then observe that the protagonist leaves the scene and then they witness that, in the absence of the protagonist, the object was transferred to a different location (e.g., a different basket). Participants’ task is typically to indicate where the protagonist will look for the object upon her return. Having developed mentalizing capabilities, one understands that the protagonist is not aware of the fact that the object was moved, and thus she would search for it in the location she originally placed it. As stated above, this requires dissociation of “my own knowledge” from what I assume other people know. And this task is clearly tapping onto a process of reasoning about a *particular* mental state, which would allow to predict a *particular* behaviour in a *specific* context. Notably, this way of operationalizing the concept of “mentalizing” shows the difference between the concept of mentalizing and the concept of the Intentional Stance: one can fail the mentalizing task (the false belief task) by attributing wrong belief to the protagonist. However, wrong belief is still a belief: one still adopts the Intentional Stance towards the protagonist (one attributes mental states to the protagonist and treats the protagonist as an agent with the capacity of having mental states), even though the attributed mental state is incorrect. Similarly, in daily lives, we might attribute wrong mental states to others (thus our process of mentalization has yielded incorrect outcome), but we still (correctly) assume that the others have mental states in general.

1.2 Intentional Stance Toward Artificial Agents

Artificial agents, such as robots, pose an interesting case for the strategy one would adopt in explaining/predicting their behaviour. On the one hand, they might look similar to humans, and behave similar to humans (as in the case of humanoids or androids), but on the other hand, they are just artefacts, thereby should not have the capacity for mental states to the same extent that other humans do (they should not be treated as intentional agents with beliefs, desires, etc.) It is supported by evidence showing that robots might not naturally evoke the adoption of the Intentional Stance to the same extent as other humans do [7]. Specifically, brain regions involved in mentalizing, namely the medial prefrontal cortex and the right temporoparietal junction [8], have been shown to be recruited only when people believed to interact with another human, but not with artificial agents, indicating that people did not involve the process of mentalizing when observing robots [9].

On the other hand, robots can be viewed as having “self-directed mechanical minds dwelling inside human-like bodies” [10, 11], and other evidence shows that humans attribute some degree of intentionality to robots. For instance, Thellman and colleagues presented a series of images and verbal descriptions of various behaviors displayed by a human or by a humanoid robot and asked participants to rate the behaviors in terms of intentionality, desirability, and controllability [12]. The authors found that participants’ interpretations of the behaviors as intentional were similar between humans and robots. The authors also showed that, when interacting with a non-anthropomorphic robot, the perceived human-likeness of its behaviors increased the likelihood of adopting the Intentional Stance [12].

Recently, Marchesi and colleagues [13] also showed that people adopt the Intentional Stance towards humanoid robots to some extent. The authors developed a tool – the InStance Test – to quantify whether a person is likely to adopt the Intentional Stance towards humanoid robots. The test consists of 34 fictional scenarios, in which the humanoid iCub robot [14] is depicted performing various daily activities. Each scenario comprises three pictures showing a sequence of events, with a scale (ranging from 0 to 100) providing a mechanistic description of that scenario on one boundary and a mentalistic description on the other. Participants’ task is to rate whether they think that the robot’s behavior is motivated by a mechanical cause (such as malfunctioning or calibration) or by a mentalistic one (such as desire or curiosity). The higher the IST score, the more likely participants were to adopt the Intentional Stance. Results showed that participants adopted the Intentional Stance to a certain degree and that individual biases may occur in the likelihood of adopting the Intentional Stance. In a further study [15], participants’ response times were measured when choosing a response option (i.e., mentalistic vs. mechanistic description) during the InStance Test with both a human and a humanoid robot. Results showed that participants were more likely to use mentalistic descriptions for the human and mechanistic descriptions for the robot. However, when looking at participants’ reaction times when giving a response, no differences emerged between the mentalistic and the mechanistic description for scenarios depicting the humanoid robot agent, suggesting that both stances (Intentional vs. Design Stance[1]) were "equally likely" to explain the behavior of the robot [15].

Overall, literature suggests that humans might, to some extent, adopt the Intentional Stance towards robots (especially those that have a human-like shape or behavior). For social robotics, it is an important issue, as the community needs to understand the conditions and factors influencing the likelihood of adopting the Intentional Stance towards robots. Adopting the Intentional Stance is likely to increase the level of social engagement and attunement, as robots that are treated as intentional agents would be perceived more as "like us" – more of social partners for our daily activities. It is quite plausible that, for example, a robot designed to remind an elderly person about taking a medication would be more successful in this task, relative to a robot perceived as a mechanical device, an automatic "alarm clock" which is simple to ignore. In such a context, it might be beneficial and useful to design a robot behavior which is likely to evoke the adoption of the Intentional Stance. On the other hand, in factory settings, where the user should focus on their individual performance, and be as efficient as possible, a social agent (with attributed intentionality) might be too distracting. Also, there might be contexts in which it is not desirable to evoke the Intentional Stance in order to not evoke over-attachment to the robot companion. In all these examples, however, one also needs to understand the specific profile of a given user. For some users, certain behaviors will easily evoke the Intentional Stance, for others, less so. Thus, in examining the factors that are crucial for evoking adoption of the Intentional Stance, one should not forget the individual characteristics of users.

1.3 Experience with Technology and Intentional Stance Toward Artificial Agents

One factor potentially playing a critical role in the adoption of the Intentional Stance towards robots is the degree of previous experience with robots. Recent findings demonstrated that it was associated with more negative attitudes towards robots and a lower tendency to perceive robots as social agents [16]. Notably, exposing participants to a longer duration of repetitive interactions with a robot also decreases their likelihood of adopting the Intentional Stance towards robots [17].

A crucial factor that is related to previous experience with robots is education. Interestingly, years of education seem to be negatively correlated with the adoption of the Intentional Stance towards robots, in such a way that the less years that individuals participated in formal education, the more likely they were to adopt the Intentional Stance [18]. The authors [18] suggested that individuals with lower education might have been less exposed to technological knowledge. This would increase the likelihood of adopting the Intentional Stance because when humans are exposed to an unknown (or not easily understandable) system, such as a robot, they are more prone to adopt the "intentional" strategy as the most efficient and familiar model of reasoning about others' behaviors [15, 19].

Following this logic, not only the level but also the type of education might play a role in the adoption of the Intentional Stance. The idea is that familiarity with robots and understanding the inner workings of robots might prevent individuals who have expertise in robotics from adopting the Intentional Stance. Conversely, those who do not have experience in robotics, but rather have being trained to extensively use their mentalizing abilities, should have an higher tendency to use the intentional strategy to explain and predict behaviors of robots, as this is the strategy in reasoning about others which they should be most familiar with.

In line with this reasoning, we set out to test whether the type of education (robotics vs. psychotherapy) affects the likelihood of adopting the Intentional Stance towards robots.

2 Aims

The present study aimed to investigate whether, and how, participants' type of education modulates their likelihood of adopting the Intentional Stance [2–4] towards robots.

To this aim, we designed a study in which we asked participants to complete the InStance Test (IST) [13], depicting the humanoid iCub robot [14] in various daily activities (see Fig. 1 for an example scenario of the IST).

To test the role of participants' type of education in the adoption of the Intentional Stance towards robots, we recruited two samples of participants varying in their type of formal education but with similar levels (years) of education. On the one hand, we recruited a sample of participants working in the robotics field, as we reasoned that their formal education led them to acquire technical knowledge about robots in terms of design, programming, and functionalities. On the other hand, we recruited a sample of psychotherapists, as we reasoned that, given their formal education, they did not have previous knowledge about robots. Conversely,

[1] Dennett [2–5] contrasts the Intentional Stance with – what he calls – the "Design" stance, namely a stance/strategy of predicting/explaining behaviors of an entity with reference NOT to mental states, but rather with how the entity was designed to behave (think of a car, coffee machine, etc.)

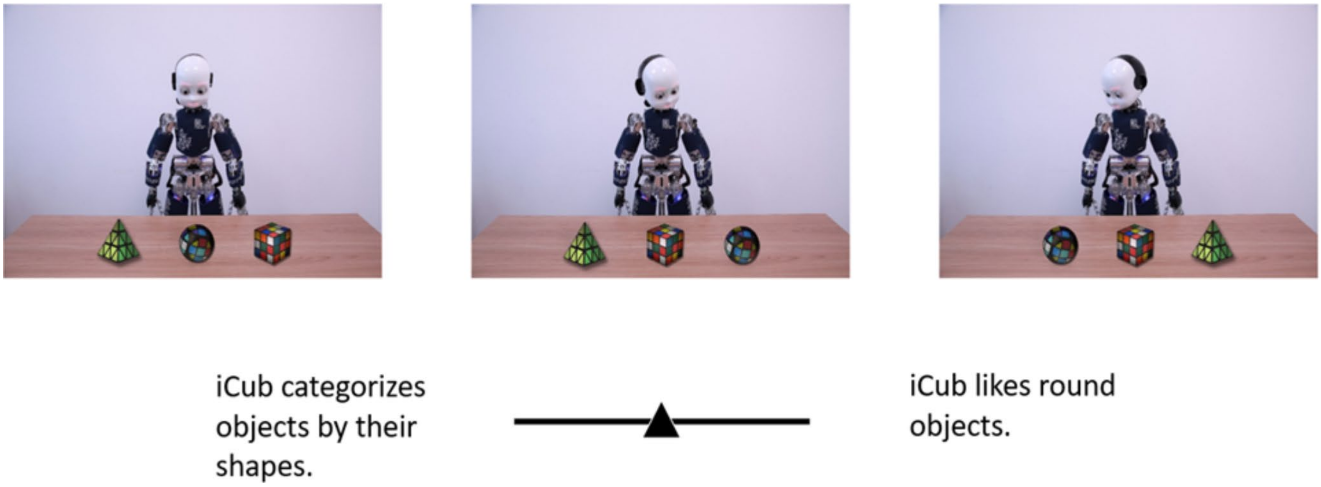

iCub categorizes objects by their shapes.
iCub likes round objects.

Fig. 1 Screenshot of the example scenario of the InStance Test (IST). In our study, the IST was administered in Italian, as well as in the original study of Marchesi and colleagues [13]

due to their professional career, we reasoned that they were well-trained in using mentalizing abilities, which might also influence their likelihood of adopting the Intentional Stance. Importantly, we designed the two samples to have a comparable duration of formal education, and hence the choice of roboticists with a Ph.D. degree or enrolled in a Ph.D. program and psychotherapists with completed certification of their formal education in a psychotherapy school (3–4 years after the master degree) or enrolled in such a program.

We assessed individuals' likelihood of adopting the Intentional Stance towards robots through participants' scores at the InStance Test (IST), while we recorded their neural activity during resting state via electroencephalography (EEG). Resting state activity has been thought to measure default neural activity when participants are not involved in completion of any task but are instructed to rest and let their minds freely wander. Specifically, the "default mode network" (DMN), a neural network strongly activated at rest, seems to be involved in social cognition in general [20–23], as well as in the adoption of the Intentional Stance [24]. In a recent study, Bossi and colleagues measured participants' resting state EEG activity, before asking them to complete the IST and measure their likelihood of adopting the Intentional Stance towards robots [24].

The authors focused on the resting state activity in the beta band, whose frequency has been correlated with the activation of cortical regions involved in the DMN [24]. They observed that it was possible to discriminate participants who were more likely to adopt the Intentional Stance towards robots from those who were more likely to adopt the Design Stance, showing that individuals' attitudes in adopting the Intentional Stance can be detected at the neural level in the resting state EEG signal.

Based on these results [24], we decided to focus on participants' resting state activity in the beta band. We also focused on participants' gamma activity, as enhanced neural oscillations in the gamma frequency range have been associated with mentalizing [25].

We hypothesized that, if participants' prior knowledge about robots, operationalized as type of education, modulates their adoption of the Intentional Stance, then we should observe a difference between two groups of participants with different education type (robotics vs. psychotherapy) in both IST scores and neural activity in the resting state (H1). Our hypothesis H1 would be tested against the null hypothesis H0, according to which participants' type of education (and thus prior knowledge about robots) does not affect their adoption of the Intentional Stance or neural activity. Thus under H0, both IST and power in beta and gamma frequency bands should be equal between therapists and roboticists. In addition, we had a more directional hypothesis (H2) according to which we expected higher IST scores, and higher power in both beta and gamma frequency bands, for therapists- whose formal training was meant to promote mentalizing skills-, as compared to roboticists (H2).

3 Materials and Methods

Participants. Two samples of participants were recruited to take part in this study. The first sample comprised right-handed participants working in the robotics field, namely those holding a Ph.D. or currently enrolled in a Ph.D. program in robotics ("Roboticists" sample; N = 16, 12 males, 4 females; Age range: 24–30 years old; M $_{\text{Age}}$ = 28.13, SD $_{\text{Age}}$ = 3.46). The second sample comprised right-handed participants working as psychotherapists, namely individuals who already finished psychotherapy school or are currently enrolled in it (“Therapists” sample; N = 17, 1 male; Age range: 26–48 years old; M $_{Age}$ = 34.18, SD $_{Age}$ = 6.19). Due to sex and age differences across the two samples, we ran additional analyses to examine whether they contributed to the effects of interest (see Supplementary Materials for more information). All participants had normal or corrected-to-normal vision and gave written informed consent before the beginning of the experiment. Our exclusion criteria comprised no history of neurological or psychiatric diseases.

The “Roboticists” sample was recruited among the employees of the Italian Institute of Technology (Genoa, Italy), and part of their salary was compensated for participation- in addition to extra holiday hours. The “Therapists” sample was recruited among the members of the Ligurian Psychologists Association (“Ordine degli Psicologi Regione Liguria”-OPLi), who received an honorarium of 50 euros for participation. All participants were naïve to the purpose of the study, and they were debriefed at the end of the experimental session.

The study has been approved by the local ethical committee (Comitato Etico Regione Liguria) and conducted following the ethical standards laid down in the 2013 Declaration of Helsinki.

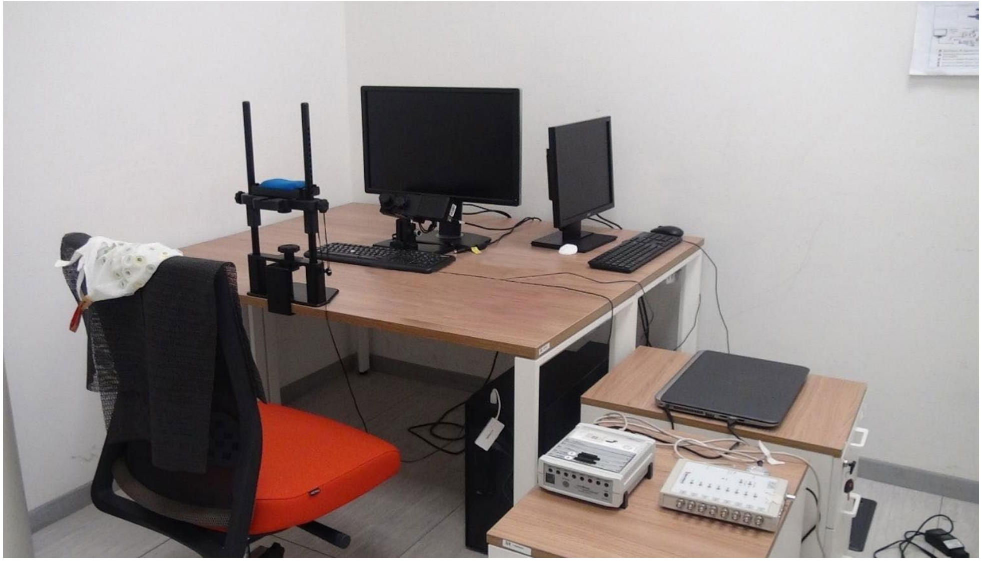
Apparatus and Stimuli. The experimental apparatus comprised a workstation equipped with a 21 inches screen to display the task (resolution: 1920 × 1080), one QWERTY keyboard and one mouse to give responses to IST, a chin rest to keep the participants’ heads as stable as possible, one laptop for the EEG recording, and one set of earphones to present the IST sentences auditorily (see [24] for a similar procedure) (Fig. 2).

Regarding the EEG apparatus, EEG data were recorded using Ag-AgCL electrodes from a 64 electrodes system, following the International 10–20 layout (ActiCap, Brain Products, GmbH, Munich, Germany). This system, which used active electrodes, is considered state-of-the-art in EEG research, and has been used in numerous studies investigating EEG activity, including in the gamma frequency range (e.g. [24, 26],). The data were referenced online to FCz. The EEG signal was amplified with a BrainAmp amplifier (Brain Products, GmbH), digitized at a 500- Hz sampling rate, and recorded. No filters were applied during EEG signal recording. Where possible, electrode impedances were kept below 10 kilo ohms for the entire duration of the experiment. The InStance Test (IST) [13] was programmed and presented using Experiment Builder Version 2.2.245. The resting state procedure was presented via Psychopy v2020.1.3 [27].

Procedure. Participants sat at approximately 80 cm from the screen. Before the IST, we recorded participants’ neural activity via EEG during a resting state session, which took five minutes. It comprised six alternating sessions of eyes-open and eyes-closed, each of them lasting 30 s. During eyes-open sessions, participants were instructed to keep their gaze on a fixation dot presented in the center of the screen. They were instructed to relax and avoid moving or blinking as much as they could. During eyes-closed sessions, they were asked to avoid movements and to wait for a beep signaling the end of the session.

After the resting state session, participants were instructed to complete the IST, in an adapted version where the response options were presented auditorily (as in [24]). They were first presented with five practice trials, in which the same scenario was always displayed to let participants familiarize themselves with the task. This scenario was not considered part of the IST. Then, participants completed the IST where they were presented with 34 scenarios in random order. Each scenario was accompanied by two descriptions presented auditorily, one of which used mentalistic and the other mechanistic vocabulary. Participants were asked

Fig. 2 Experimental setup



to respond to the scenarios by moving the cursor towards one of the two extremes, where one extreme represented the mechanistic description, and the other the intentional description. One of the extremes of the slider reads “A” while the other, “B”, refers to the description A/B. The association between mechanistic and mentalistic statements with descriptions A and B was counterbalanced across trials.

As the present study focuses on resting-state EEG activity, and on participants’ scores at the IST test, here we do not describe the trial sequence of the IST in detail (however, see Fig. 3 for an example of an experimental trial). Information about the details of the procedure used in the IST trial can be found in Bossi and colleagues’ paper [24].

4 Data Processing

For the analysis of the IST scores, we considered only participants who fully completed the IST (i.e., all 34 trials). Therefore, from the “Roboticists” sample we excluded 4 participants, resulting in a final sample size of N = 12 (9 males, 3 females). From the “Therapists” sample, we excluded 3 participants, resulting in a final sample size of N = 14 (all females). For the EEG analysis, as we focused on pre-task resting state activity, we did not exclude participants who did not complete the IST. However, we excluded two participants for reasons of data quality; the ICA decompositions for these data resulted in too many noisy components (greater than 70% of the components appeared to represent non-brain signals). One participant was excluded from the “Roboticists” sample, resulting in a final sample size of N = 15 (11 males, 4 females), and one participant was excluded from the “Therapists” sample, resulting in a final sample size of N = 16 (all females).

EEG data. Resting state EEG data were preprocessed and analyzed using MATLAB version R2020b (The MathWorks Inc., 2020), as well as EEGLab [28] and FieldTrip toolboxes [29], along with customized scripts and R Studio [30]. Data were down-sampled to 250 Hz, band pass filtered between 0.5 and 100 Hz, and notch filtered at 50 Hz to remove line noise. Then, data were segmented into pseudo epochs of 1000 ms, to make the subsequent processing steps easier. Epochs with prominent artifacts (e.g., muscle noise) were removed through visual inspection, as were bad channels. On average, we removed 84.97 pseudo epochs per participant (SD = 25.92) and 3.1 channels per participant (SD = 1.99). Following the visual inspection, the data were re-referenced to the average of all electrodes. Independent component analysis (ICA) was applied to the data to further remove artifacts related to eye blinks or eye movements, and other remaining artifacts. On average, we removed 32.80 ICs per participant (SD = 5.87). Following artifacts’ removal via ICA, the removed channels were spatially interpolated, and the data were again re-referenced to the average of all electrodes. Then, data were separated into eyes-open and eyes-closed segments. Based on previous work [24], eyes-open segments were analyzed using a Fast Fourier Transform (FFT), with Hanning tapers. Frequencies from 2 to 60 Hz were used in the FFT, in steps of 1 Hz.

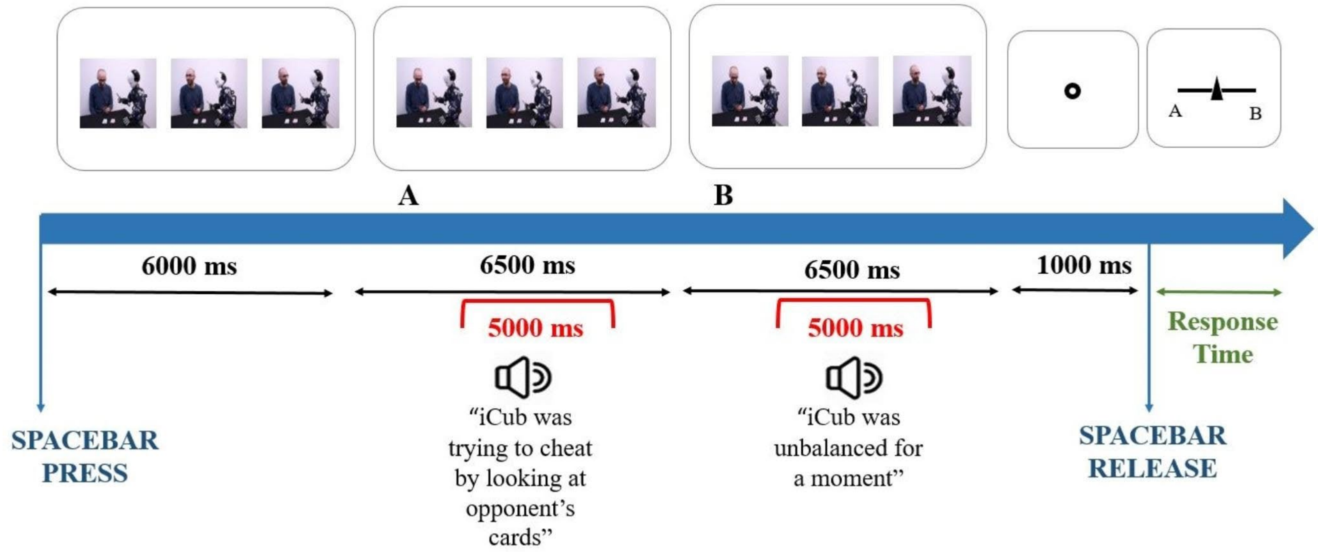

A
B
A
B
6000 ms
6500 ms
6500 ms
1000 ms
5000 ms
5000 ms
“iCub was trying to cheat by looking at opponent’s cards”
“iCub was unbalanced for a moment”
Response Time
SPACEBAR PRESS
SPACEBAR RELEASE

Fig. 3 Example of an experimental trial. Participants started the trial by pressing the spacebar, and they had to keep the spacebar pressed until they decided which response to give. They heard both response options while the scenario was displayed on the screen; the order of response options was counterbalanced across participants. After the presentation of the second sentence, participants were presented with a fixation dot; then, upon the presentation of the sliding scale with letters “A” and “B” on the two extremes, they decided which response to give and they reached the mouse as fast as possible to move the cursor towards the selected option

5 Statistical Analysis

IST scores. Our aim here was to assess whether participants' type of education modulates the adoption of the Intentional Stance in terms of IST scores. According to our initial hypothesis (H1), we should observe a difference in IST scores between the group of therapists and the group of roboticists. According to our second, more directional hypothesis (H2), we should observe higher IST scores for therapists, as compared to roboticists.

First, IST scores were calculated by converting the point on the line where participants placed the slider into a 0-100 scale (where 0 would be the most extreme "mechanistic" point on the line, and 100 would be the most extreme "intentional" point) for each item. Then, we averaged scores across the items, per participant. Participants' mean IST score corresponds to their bias, i.e., if they are more likely to adopt the Intentional or the Design Stance. In other words, the higher the score, the more participants were prone to adopt the Intentional Stance.

Then, we checked whether the data met the assumption of normality, which was assessed through the Shapiro-Wilk test. Results showed that both samples of data were normally distributed ("Roboticists" sample, $p = 0.67$; "Therapists" sample, $p = 0.93$). Then, with participants' mean IST scores as the dependent variable, we compared the two samples (roboticists vs. therapists) through a Bayesian t-test using the *Bolstad* package [31] in R Studio v.4.0.5 [30]. The t-statistic, with the associated p-value and the corresponding 95% confidence intervals (CI), is reported below.

Furthermore, we performed also a t-test using the Monte Carlo method. It is a simulation technique in which specific selected properties of a sample are computer-generated to assess the behavior of a statistical procedure or parameter under varying conditions (https://dictionary.apa.org/monte-carlo-research). In the context of this study, both "Roboticists" and "Therapists" samples were small in size ($N_{\text{Roboticists}} = 12$; $N_{\text{Therapists}} = 14$), and presented both gender and age disparities as reported above (see *Participants* section in Materials and Methods). Therefore, we used Monte Carlo simulations to run k-independent comparisons (where $k = 10{,}000$) between the two samples. It was made to see whether the results obtained by comparing participants' mean IST scores across the two samples were robust enough to test our initial H1 hypothesis- that is, a difference in the IST scores between the two groups. Indeed, Monte Carlo simulations allowed us to assess the power of the comparisons, i.e., the proportion of significant p-values (threshold: $p < 0.05$) reflecting the likelihood of detecting a significant difference in the IST scores between the two samples when there is one. Notably, the value of the power ranges between 0 and 1, where 0 denotes that no significant comparisons ($p < 0.05$) emerged, and 1 denotes that all k comparisons were statistically significant.

By using the corresponding Mean and SD for each sample, we ran 10,000 independent comparisons, each of those resulting in the computation of a p-value showing whether there was a difference between the two samples. Then, we estimated the proportion of instances of $p < 0.05$ from the distribution of the simulated p-values, which gives a measure of the power.

EEG data. Our aim here was to assess whether participants' type of education modulates the adoption of the Intentional Stance at the neural level, i.e., in terms of neural activity during resting state. According to our directional hypothesis (H2), we should observe higher power in beta and gamma frequency bands, associated with mentalizing and adoption of the Intentional Stance, in therapists, whose formal training was meant to promote mentalizing skills, as compared to roboticists.

To compare the sensor-level EEG activity in the beta (frequency range: 12–27 Hz) and gamma range (frequency range: 28–45 Hz) between the two groups (i.e., "Roboticists" vs. "Therapists" samples), cluster-based non-parametric permutation analyses were performed using a Monte Carlo method based on paired t-statistics. Using data averaged across the entire frequency of interest, t-values were selected with a threshold of $p < 0.05$, and clustered based on neighboring electrodes. Subsequently, cluster-level statistics were calculated by summing the t-values in each cluster. Subsequent comparisons were performed for the maximum values of summed t-values. Using a random partition-based permutation test (number of permutations = 1,500), the hypothetical null distribution of the maximum of summed cluster level t-statistics was obtained. The actual cluster-level statistics extracted from the data were then compared to the null distribution. The significance level was set to 0.05. Thus, cluster-level statistics from the actual data were considered significant if they were larger than 95% of the cluster-level statistics in the null distribution.

Demographics, descriptive statistics as well as additional analyses, such as comparisons with a sample drawn from the General Population collected by Bossi and colleagues [24], sex, and age analyses can be found in Supplementary Materials.

6 Results

IST scores. Results of the Bayesian t-test showed a significant difference between the two groups [$t_{(24)} = -2.45$, $p = 0.02$, 95% CI = (-28.1; -1.42)], with higher IST scores for therapists compared to roboticists (Mean $\text{IST}_{\text{Therapists}} = 51.9$; Mean $\text{IST}_{\text{Roboticists}} = 36.7$) (see Fig. 4).

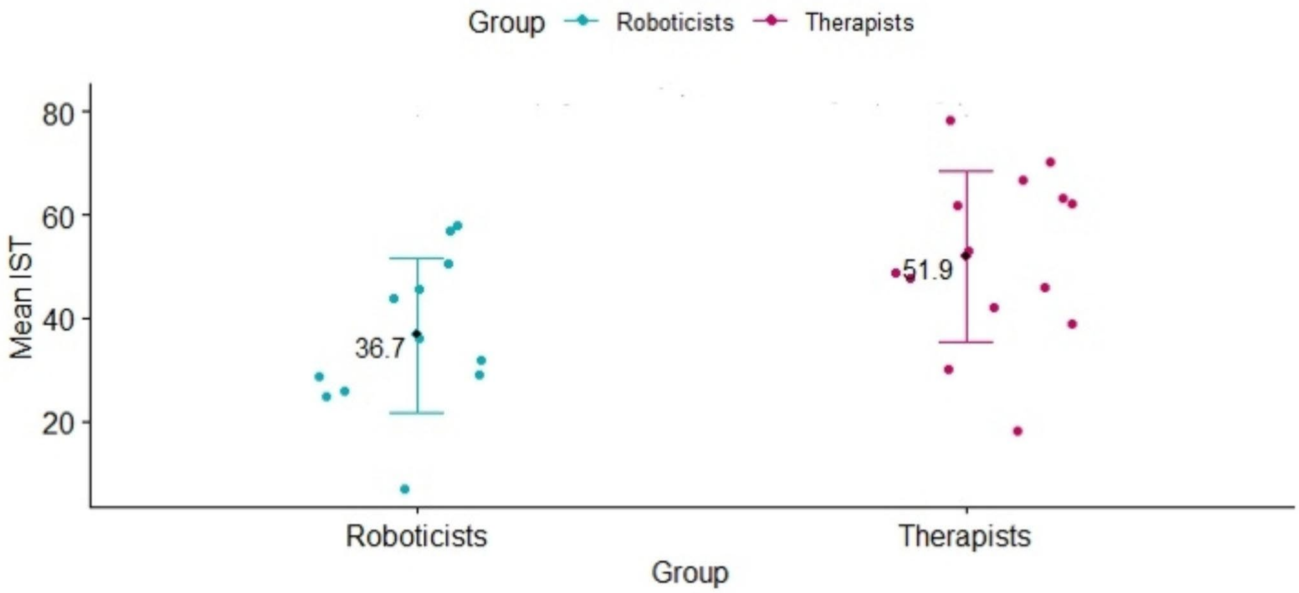


Fig. 4 Participants' mean at the IST, plotted as a function of Group (Roboticists vs. Therapists). Error bars represent the standard deviation (SD)

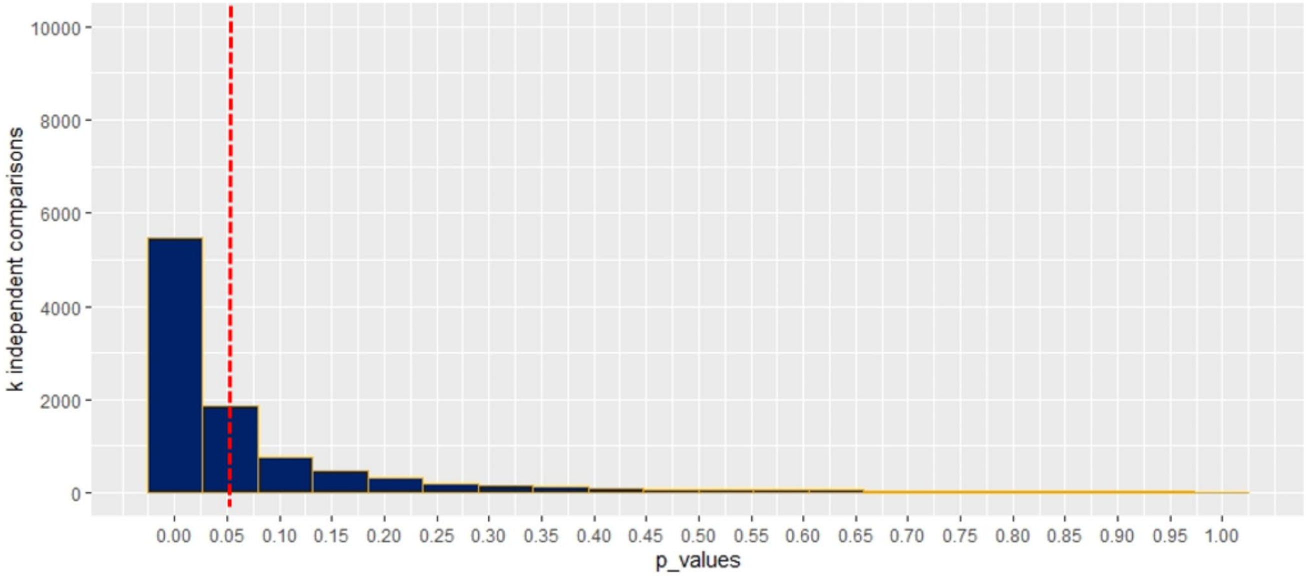


Fig. 5 Distribution of the p-values (on the *x*-axis) resulting from k = 10,000 independent comparisons (on the *y*-axis) between the IST scores of roboticists and therapists, using the Monte Carlo method. The red dashed line represents the threshold of significance ($p < 0.05$). The higher the number of instances of significant p-values (i.e., the ones on the left side of the red dashed line), the higher the power of the comparisons

Results of the t-test using Monte Carlo simulation showed that, by running 10,000 independent comparisons between the two groups, the total number of significant ($p < 0.05$) comparisons was equal to 6580, resulting in a power of 0.66. It would mean that, by comparing the two samples in 10,000 independent simulations, 66% of these comparisons showed a significant difference in the IST scores between the two (see Fig. 5).

EEG data (Resting State). The cluster-based permutation analysis did not show any clusters in which beta power differed significantly between the groups, indicating no significant difference in beta range activity between the two groups. However, the analysis revealed a significant difference in gamma range activity between the groups, as therapists showed higher power in the gamma range than roboticists, in a posterior cluster ($p = 0.03$, corrected for multiple comparisons). The cluster consisted of 5 right lateralized electrodes (O2, PO4, PO8, P6, P8) (see Fig. 6).

7 Discussion

The present study aimed to investigate whether participants' type of education modulated their likelihood of adopting the Intentional Stance [2–4] towards robots. To this aim, we recruited two samples of participants with different backgrounds, in terms of formal education. The first sample comprised participants working in the field of robotics, thus having prior knowledge about the functionality and inner workings of robots (the "Roboticists" sample); the second sample comprised participants working as psychotherapists, with no previous knowledge about robots, but well trained in mentalizing abilities due to their professional career (the "Therapists" sample).

We employed the InStance Test (IST) [13] to measure participants' likelihood of adopting the Intentional Stance toward robots. We also assessed individuals' tendency to mentalizing at the electrophysiological level, as participants' bias to adopt the Intentional Stance might be detected at the neural level, in EEG resting state activity [24].

Results showed a significant difference in the IST scores between the two groups ("Roboticists" vs. "Therapists"). Specifically, therapists scored higher in the IST compared to roboticists, thus displaying a greater likelihood of adopting the Intentional Stance towards robots. Moreover, this result was mirrored by participants' neural activity during the resting state. EEG results showed that therapists displayed higher power in the gamma frequency range compared to roboticists. This difference emerged in a right lateralized occipital-partial cluster of electrodes. Enhanced gamma activity has been associated with mentalizing and Intentional Stance [24, 25]. More specifically, the predominantly right-lateralized topography in which we find differences in gamma power in our study is consistent with the activity in the right temporal parietal junction (rTPJ), which has been implicated in mentalizing processes [25, 32, 33] as well as the adoption of the Intentional Stance [9, 24]. Therefore, one possible explanation might be that therapists, as compared to roboticists, displayed higher power in gamma frequency bands due to their higher mentalizing abilities, resulting from their education. There were no differences in beta activity across the two groups, in contrast to the resting state results of Bossi et al. [24], whose study was conducted on a sample from the general population, and not experts.

Taken together, our results showed that participants with prior knowledge about robots (i.e., roboticists) tended to attribute less intentionality to robots compared to participants without prior technical knowledge about robots (i.e., therapists). This is in line with previous literature, according to which the less educated people are (and presumably less knowledge they have about technology), the more likely they are to adopt Intentional Stance toward technological entities [18]. Indeed, being less informed about how a system – in this case, the iCub robot – has been designed and works would lead people to treat it as an intentional system, as the Intentional Stance seems to be the most available and default predictive strategy to interpret and explain the behavior of systems that resemble humans in physical appearance [18]. Conversely, it may be that the more people have been exposed to robots, the more they acquire knowledge about the way a system is designed and programmed to behave (as in the case of people working in the robotics field). Consequently, it may encourage such experts to adopt the Design Stance towards robots, rather than the Intentional Stance [16].

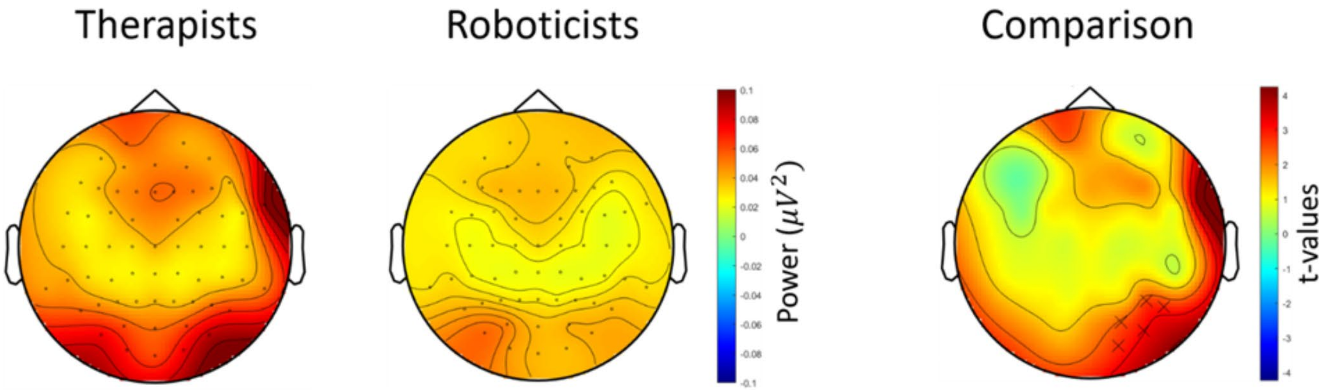


Fig. 6 Results from the analyses of resting-state gamma activity. The first two topography plots show the average gamma range (28–45 Hz) power, averaged across all subjects for the therapists (left) and roboticists (right). The gamma band power averaged across the entire frequency range was calculated by applying an FFT to the entire duration of the eyes-open resting state data. The third topography shows the t-value map of the cluster of statistically significant differences in gamma range activity between the two groups, calculated using cluster-based non-parametric permutation analysis. Channels showing statistically significant differences are marked by an X

In more general terms, these results extend current knowledge on factors influencing the likelihood of adopting the Intentional Stance towards robots. Past research showed that implementing human-like characteristics in robots, in terms of physical appearance or behavior, facilitates social interactions with them (e.g. [34, 35],), leading people to display more positive attitudes towards them, for example in terms of pleasantness [36] and acceptability [37]. More recent evidence showed that even subtle cues such as the range of behavioral variability displayed by the robot in a joint action task with a human can be considered a "hint of humanness" that people can use to ascribe human-like features to non-human agents such as robots [38]. Our results showed that not only robot's characteristics, but also humans' individual differences (such as type of formal education) contribute to the likelihood of adopting the Intentional Stance towards robots. Bossi and colleagues [24] were the first to demonstrate that it is possible to differentiate neural activity between people who adopt Intentional Stance towards a humanoid robot and those who adopt the Design Stance. Our study took a further step ahead, showing that neural activity, recorded with the EEG, can also differentiate between different participants' educational background, which then translates to different likelihoods of adopting the Intentional Stance towards robots. This result has an important consequence not only theoretically – by showing that individual differences, such as type of education, can relate with differential activity of the brain – but also practically. One day in the future, perhaps it will be possible to design robots that will be adapting their behavior to the specific profile of the user. For some users the robots would behave in such a way as to evoke higher likelihood of adopting the Intentional Stance, while for others, it would behave more mechanistically.

Overall, from the perspective of social robotics, our study highlights that successful integration of robots into human (social) lives and environment does not rely only on the technological capabilities of the robots, but also on the profile and individual characteristics of the human users. In other words, designing and developing social robots that can spontaneously evoke the adoption of the Intentional Stance based on certain characteristics of the human users (i.e., type of prior knowledge about robots) might be beneficial. For example, in healthcare settings, a robot that can socially attune with elderly, and thus create engaging and positive interactions with them, might improve their mood, decrease their feelings of loneliness, and strengthen connections with others, with positive consequences on the elderly's quality of life [39]. In another relevant context such as education, proper interventions focused on equipping individuals with a certain type of knowledge- both in terms of years and type of education- might promote (or, quite the opposite, avoid) certain attitudes. In some contexts, it might be beneficial that the user treats the robot as an intentional agent, in other contexts, it might be crucial that the robot is treated completely mechanistically.

Ideally, a social robot should be able to recognize the needs of the humans, and then respond adequately by displaying the proper level of engagement and social attunement. Importantly, our study's main contribution to the field of social robotics is the finding that it is possible to detect (at least some) individual characteristics of the user based on neural activity. This might be an efficient way to design robots that would receive such neural signals and adapt their behavior accordingly.

However, having said this, it is important to note that endowing robots with adaptive (and thus, extremely social) behaviors might have undesirable consequences. For example, the risk of experiencing an excessive emotional attachment to the robot might make people with psychosocial risks, loneliness, depression, or social anxiety even more vulnerable. These aspects need to be discussed from an ethics point of view, at the policymaking level.

Acknowledgements The authors would like to thank Ordine degli Psicologi- Regione Liguria for their availability to recruit participants among their professionals. They would also thank all the workers within Italian Institute of Technology (IIT) based in Genoa, Italy for kindly volunteering to participate in the study.

Author Contribution CR collected and analyzed the behavioral data, discussed and interpreted the results, and wrote the manuscript. UPN analyzed the EEG data, discussed and interpreted the results, and wrote the manuscript. FC designed and programmed the experiment, discussed and interpreted the results, and wrote the manuscript. AW designed the experiment, discussed and interpreted the results, and wrote the manuscript. All the authors revised the manuscript.

Funding Open access funding provided by Istituto Italiano di Tecnologia within the CRUI-CARE Agreement. This work has received support from the European Research Council under the European Union's Horizon 2020 research and innovation program, ERC Starting Grant, G.A. number: ERC-2016-StG-715058, awarded to AW. The content of this article is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.
Open access funding provided by Istituto Italiano di Tecnologia within the CRUI-CARE Agreement.

Data Availability The datasets analyzed during the current study are available at the following link: https://osf.io/kxf7u/ (OSF project, name: "How education type affects individuals' adoption of Intentional Stance towards robots: an EEG study").

Declarations

Competing Interests The authors declare that this study has been conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

References

1. Luyten P, Fonagy P (2015) The neurobiology of mentalizing. Personality Disorders: Theory Research and Treatment 6(4):366. https://doi.org/10.1037/per0000117
2. Dennett DC (1971) Intentional systems. J Philos 68. https://doi.org/10.2307/2025382
3. Dennett DC (1987) The intentional stance. MIT Press
4. Dennett DC (2009) Intentional Systems Theory. In The Oxford Handbook of Philosophy of Mind. https://doi.org/10.1093/oxfordhb/9780199262618.003.0020
5. Perez-Osorio J, Wykowska A (2020) Adopting the intentional stance toward natural and artificial agents. Philosophical Psychol 33(3):369–395. https://doi.org/10.1080/09515089.2019.1688778
6. Wimmer H, Perner J (1983) Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's. Underst Decept Cognition 13(1):103–128. https://doi.org/10.1016/0010-0277(83)90004-5
7. Krach S, Hegel F, Wrede B, Sagerer G, Binkofski F, Kircher T (2008) Can machines think? Interaction and perspective taking with robots investigated via fMRI. PLoS ONE. https://doi.org/10.1371/journal.pone.0002597
8. Frith CD, Frith U (1999) Interacting minds–a biological basis. Sci (New York N Y) 286(5445):1692–1695. https://doi.org/10.1126/science.286.5445.1692
9. Chaminade T, Rosset D, Da Fonseca D, Nazarian B, Lutcher E, Cheng G, Deruelle C (2012) How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. Front Hum Neurosci 6:103. https://doi.org/10.3389/fnhum.2012.00103
10. Thellman S, de Graaf M, Ziemke T (2022) Mental State Attribution to Robots: a systematic review of conceptions, methods, and findings. ACM Trans Human-Robot Interact (THRI) 11(4):1–51. https://doi.org/10.1145/3526112
11. Brink KA, Gray K, Wellman HM (2019) Creepiness creeps in: uncanny valley feelings are acquired in childhood. Child Dev 90(4):1202–1214. https://doi.org/10.1111/cdev.12999
12. Thellman S, Silvervarg A, Ziemke T (2017) Folk-psychological interpretation of human vs. humanoid robot behavior: exploring the intentional stance toward robots. Front Psychol 8:1962. https://doi.org/10.3389/fpsyg.2017.01962
13. Marchesi S, Ghiglino D, Ciardo F, Perez-Osorio J, Baykara E, Wykowska A (2019) Do we adopt the intentional stance toward humanoid robots? Front Psychol 10:450. https://doi.org/10.3389/fpsyg.2019.00450
14. Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., … & Montesano, L. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. Neural networks, 23(8–9), 1125–1134. https://doi.org/10.1016/j.neunet.2010.08.010
15. Marchesi S, Spatola N, Perez-Osorio J, Wykowska A (2021), March Human vs Humanoid. A behavioral investigation of the individual tendency to adopt the intentional stance. In Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 332–340. https://doi.org/10.1145/3434073.3444663
16. Ciardo F, Ghiglino D, Roselli C, Wykowska A (2020) The effect of Individual Differences and Repetitive Interactions on Explicit and Implicit attitudes towards Robots. In:, et al. Social Robotics. ICSR 2020. Lecture Notes in Computer Science, vol 12483. Springer, Cham. https://doi.org/10.1007/978-3-030-62056-1_39
17. Abubshait A, Wykowska A (2020) Repetitive robot behavior impacts perception of intentionality and gaze-related attentional orienting. Front Rob AI 7:565825. https://doi.org/10.3389/frobt.2020.565825
18. Ghiglino D, Wykowska A (2020) When robots (pretend to) think. Artificial Intelligence. Brill mentis, pp 49–74
19. Spatola N, Marchesi S, Wykowska A (2022) Cognitive load affects early processes involved in mentalizing robot behaviour. Sci Rep 12(1):1–14. https://doi.org/10.1038/s41598-022-19213-5
20. Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN (2007) Wandering minds: the default network and stimulus-independent thought. Science 315(5810):393–395. https://doi.org/10.1126/science.1131295
21. Northoff G, Heinzel A, De Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006) Self-referential processing in our brain—a meta-analysis of imaging studies on the self. NeuroImage 31(1):440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002
22. Schilbach L, Eickhoff SB, Rotarska-Jagiela A, Fink GR, Vogeley K (2008) Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the default system of the brain. Conscious Cogn 17(2):457–467. https://doi.org/10.1016/j.concog.2008.03.013
23. Xie, X., Bratec, S. M., Schmid, G., Meng, C., Doll, A., Wohlschläger, A., … Sorg, C. (2016).How do you make me feel better? Social cognitive emotion regulation and the default mode network. NeuroImage, 134, 270–280. https://doi.org/10.1016/j.neuroimage.2016.04.015
24. Bossi F, Willemse C, Cavazza J, Marchesi S, Murino V, Wykowska A (2020) The human brain reveals resting state activity patterns that are predictive of biases in attitudes toward robots. Sci Rob 5(46):eabb6652. https://doi.org/10.1126/scirobotics.abb6652
25. Cohen MX, David N, Vogeley K, Elger CE (2009) Gamma-band activity in the human superior temporal sulcus during mentalizing from nonverbal social cues. Psychophysiology 46(1):43–51. https://doi.org/10.1111/j.1469-8986.2008.00724.x
26. Wang R, Yu R, Tian Y, Wu H (2022) Individual variation in the neurophysiological representation of negative emotions in virtual reality is shaped by sociability. NeuroImage 263:119596. https://doi.org/10.1016/j.neuroimage.2022.119596
27. https://doi.org/10.3758/s13428-018-01193-y
28. Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including Independent component analysis. J Neurosci Methods 134(1):9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009
29. Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience, 2011. https://doi.org/10.1155/2011/156869
30. Team RC R: A Language and Environment for Statistical Computing. http://www.R-project.org/
31. Curran J (2020) Bolstad functions. R package. Ver. 0.2–41
32. Koster-Hale J, Richardson H, Velez N, Asaba M, Young L, Saxe R (2017) Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. NeuroImage 161:9–18. https://doi.org/10.1016/j.neuroimage.2017.08.026

33. Donaldson PH, Kirkovski M, Rinehart NJ, Enticott PG (2019) A double-blind HD-tDCS/EEG study examining right temporoparietal junction involvement in facial emotion processing. Soc Neurosci 14(6):681–696. https://doi.org/10.1080/17470919.2019.1572648
34. Fink J (2012) Anthropomorphism and human likeness in the design of Robots and Human-Robot Interaction. In: Ge SS, Khatib O, Cabibihan JJ, Simmons R, Williams MA (eds) Social Robotics. ICSR 2012. Lecture Notes in Computer Science, vol 7621. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34103-8_20
35. Fong T, Nourbakhsh I, Dautenhahn K (2003) A Survey of Socially Interactive Robots. Robot Auton Syst 42:143–166. https://doi.org/10.1016/S0921-8890(02)00372-X
36. Axelrod L, Hone K (2005) E-motional advantage: performance and satisfaction gains with affective computing. Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems, 1192-95. https://doi.org/10.1145/1056808.1056874
37. Goetz J, Kiesler S (2002), April Cooperation with a robotic assistant. In *CHI'02 Extended* Abstracts on Human Factors in Computing Systems, 578–579. https://doi.org/10.1145/506443.506492
38. Ciardo F, De Tommaso D, Wykowska A (2022) Human-like behavioral variability blurs the distinction between a human and a machine in a nonverbal turing test. Sci Rob 7(68):eabo1241. https://doi.org/10.1126/scirobotics.abo1241
39. Broekens J, Marcel H, Rosendal H (2009) Assistive social robots in elderly care: a review. Gerontechnology 8 294–103. https://doi.org/10.4017/gt.2009.08.02.002.00

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cecilia Roselli is currently a postdoctoral researcher in the Social Cognition in Human-Robot Interaction (S4HRI) unit, at the Italian Institute of Technology (Genoa, Italy). She received her Bachelor's (2015) and Master's (2017) degrees in Psychology at the University of Turin; then, she received her Ph.D. in Bioengineering and Robotics at the Università degli Studi di Genova in 2022. Her research interests primarily focus on social cognition. Her current research investigates the vicarious Sense of Agency phenomenon in a shared social context with various kinds of artificial agents, and how familiarity with robots impacts the likelihood of attributing intentionality to them.

Uma Prashant Navare has a background in neuroscience. As of 2020, she had been pursuing her Ph.D. in Computer Science at the University of Manchester (UK) and the Italian Institute of Technology (Genoa, Italy). Her research interests include understanding the behavioral and neural mechanisms involved in both human-human and human-robot interactions, focusing specifically on the various factors involved in joint actions.

Francesca Ciardo is Assistant Professor at the department of Psychology at the University of Milano-Bicocca. She has been MSCA Researcher in the group of Social Cognition in Human–Robot Interaction (S4HRI) at the Italian Institute of Technology (2021–2023). She obtained her Ph.D. in Neuroscience at the University of Modena and Reggio Emilia, Italy (2015). Her research investigates cognitive mechanisms underlying human-human and human-robot interaction. Her work focuses on attentional mechanisms underlie joint attention and joint action. Francesca Ciardo research interests are social cognition, joint attention and joint action.

Agnieszka Wykowska is the head of the unit "Social Cognition in Human-Robot Interaction" at the Italian Institute of Technology (IIT), in Genoa, Italy. She is also the Coordinator of the Center for Human Technologies, at the IIT. Her background is Cognitive Neuroscience (2006), Ludwig Maximilian University Munich (LMU) and philosophy (2001), Jagiellonian University in Krakow. She obtained a Ph.D. in psychology (2008) from the LMU. In 2016 she was awarded the ERC Starting grant "InStance: Intentional Stance for Social Attunement". She is Editor-in-Chief of International Journal of Social Robotics. Since July 2022 she serves in the role of President of the European Society for Cognitive and Affective Neuroscience (ESCAN). She is also board member of Association of ERC Grantees and a delegate to the European Research Area (ERA) Forum – an EU expert group shaping EU science policies. Her research foci are interdisciplinary, bridging psychology, cognitive neuroscience, robotics and healthcare. She combines cognitive neuroscience methods with human-robot interaction to understand the human brain mechanisms in interaction with other humans and with technology.