Research Article

Cecilia Roselli, Francesca Ciardo, and Agnieszka Wykowska*

Social inclusion of robots depends on the way a robot is presented to observers

Abstract: Research has shown that people evaluate others according to specific categories. As this phenomenon seems to transfer from human–human to human–robot interactions, in the present study we focused on (1) the degree of prior knowledge about technology, in terms of theoretical background and technical education, and (2) intentionality attribution toward robots, as factors potentially modulating individuals' tendency to perceive robots as social partners. Thus, we designed a study where we asked two samples of participants varying in their prior knowledge about technology to perform a ball-tossing game, before and after watching a video where the humanoid iCub robot was depicted either as an artificial system or as an intentional agent. Results showed that people were more prone to socially include the robot after observing iCub presented as an artificial system, regardless of their degree of prior knowledge about technology. Therefore, we suggest that the way the robot was presented, and not the prior knowledge about technology, is likely to modulate individuals' tendency to perceive the robot as a social partner.

Keywords: knowledge, technology, intentionality attribution, cyberball, human–robot interaction

* **Corresponding author: Agnieszka Wykowska,** Social Cognition in Human-Robot Interaction, Fondazione Istituto Italiano di Tecnologia, Center for Human Technologies, 16152 Genova, Italy, e-mail: Agnieszka.Wykowska@iit.it
Cecilia Roselli: Social Cognition in Human-Robot Interaction, Fondazione Istituto Italiano di Tecnologia, Center for Human Technologies, 16152 Genova, Italy; DIBRIS, Dipartimento di Informatica, Bioingegneria, Robotica ed Ingegneria dei Sistemi, 16145 Genova, Italy
Francesca Ciardo: Social Cognition in Human-Robot Interaction, Fondazione Istituto Italiano di Tecnologia, Center for Human Technologies, 16152 Genova, Italy

1 Introduction

Social categorization is a key mechanism of social cognition in humans. We tend to categorize others based on various cues, such as gender, age, and ethnicity [1]. Social categorization allows us to cope with the complexity of social information we process in everyday life [2,3], reducing the amount of novel information that needs to be processed by grouping information into a single category. Indeed, from early infancy, our brain uses various strategies to deal with the abundance of information it needs to process. One strategy to avoid overload is "chunking". Chunking of information occurs based on semantic relatedness and perceptual similarity, which are often processed and stored in our memory together to allow us to recall better and faster more information [4].

Notably, such a "chunked" processing strategy seems to be involved also in social cognition, where, for example, group members are represented as interchangeable parts of a global, heuristic whole [5]. This way, the complexity of representing each group member is overridden by the less cognitively demanding processing of chunks of information (see ref. [6] for a general discussion of the efficiency of chunked processing). Interestingly, chunking has been proposed as a potential explanation of why people are in general more able to recall individual information about minority members [7]. Indeed, compared to majority members, minority members are fewer, which implies a smaller information load because it can be "chunked" together in memory as a single unit of encoded information [3]. In other words, chunking represents a cognitive "shortcut". First, it allows for categorizing the target of perception at the group level, because the group members can be easily discerned based on cues such as sex, age, and ethnic identity. Then, only after processing available information about the target at the group level, the person-level information and the personal identity can be construed [8,9]. In shared social contexts with others, chunking simplifies perception and cognition by detecting inherent shared characteristics and imposing a structure on the social

world [10]. Consequently, people categorize themselves and others into differentiated groups (in- and out-groups). Once determined, group categorization shapes downstream evaluation and behavior, often without awareness [11].

In summary, social categorization shapes the way people interact with others, making them develop a stronger preference for people who are recognized as part of their in-group [12]. Being part of a group has numerous benefits: indeed, groups provide social support, access to important resources, protection from dangers, and the possibility to create bonds with potential mates [13,14]. Therefore, it is not surprising that group membership represents a crucial aspect of human life, which has been extensively investigated in psychological research (*e.g.*, see ref. [15] for a review).

Recently, social inclusion became a relevant topic also in the human–robot interaction (HRI) field, as evidence showed that humans adopt similar social cognitive mechanisms to robots as those adopted toward other humans [16]. For example, Eyssel and Kuchenbrandt found that human users prefer to interact with robots that are categorized as in-group members [17]. Specifically, German participants were presented with a picture of a humanoid robot, which they believed to belong either to their national in-group (Germany) or to the out-group (Turkey). When asked to rate the robot regarding its degree of anthropomorphism, warmth, and psychological closeness, participants tended to evaluate more positively the robot that was presented as an in-group member, compared to the out-group one.

A task commonly used in social psychology research to evaluate ostracism and social inclusion in a more implicit way is the Cyberball paradigm [18,19], a task in which participants believe that they are playing online a ball-tossing game with two or more partners, which are animated icons or avatars controlled by the computer program. During the task, the program can vary the degree to which the ball is tossed toward the players. For instance, ostracized players are not passed the ball after two initial tosses and thus obtain fewer ball tosses than the other players. Included players are repeatedly passed the ball and obtain an equal number of ball tosses as the other players. The Cyberball paradigm has been extensively used as an implicit measure of social inclusion in many different experimental contexts (*e.g.*, [20–23]). For example, in a previous study [24] the ethnicity of confederates was manipulated so that Caucasian American participants performed the Cyberball task with either same-ethnicity (*i.e.*, Caucasian American) or other-ethnicity confederates (*i.e.*, African American). Results showed that being included or ostracized by in-group members

intensified the experience of either exclusion or inclusion. In other words, ostracism was evaluated as more painful and social inclusion as more positive when carried out by in-group members (*i.e.*, same-ethnicity confederates) [24].

Individual differences have also been shown to affect social inclusion. For example, individual traits such as self-esteem, narcissism, and self-compassion seem to modulate people's tendency to socially include others, and thus, they should be taken into consideration when developing interventions aimed at reducing aggression in response to social exclusion [25]. Interestingly, the impact of individual differences on social inclusion does apply not only to the inclusion of other humans but also to artificial agents such as robots [26,27]. For example, age seems to play a critical role, as demonstrated by a recent study testing a group of clinicians who conducted a robot-assisted intervention [23]. Specifically, when investigating individual differences in both explicit and implicit attitudes toward robots, it emerged that older clinicians displayed more negative attitudes. Moreover, also the level of education has been shown to modulate the social inclusion of robots, in such a way that the more educated people were, the less they were prone to perceive robots as social entities [28]. Individual differences in the social inclusion of robots are also driven by culture, leading people to express different levels of trust, likeability, and engagement toward robots [29]. In a recent study [30], participants of two different nationalities, *i.e.*, Chinese and UK participants, performed a modified version of the Cyberball task (*e.g.*, [18,19,23]), to assess their tendency to socially include the humanoid robot iCub [31] in the game. Interestingly, results showed that only cultural differences at the individual level, but not at the country level, were predictive of the social inclusion of robots [30]. In other words, the more individual participants displayed a collectivistic stance, the more they tended to socially include the robot in the Cyberball game. However, social inclusion was not affected by participants' nationality, namely whether they belong to a collectivistic (*i.e.*, Chinese) rather than an individualistic culture (*i.e.*, UK) [30].

Social inclusion and exclusion are related to prior knowledge or biases that people have toward others. Indeed, it has been demonstrated that, when people are repeatedly presented with novel stimuli, they tend to develop a preference for them, as repeated exposure allows people to gain knowledge about them. This psychological phenomenon has been called the *mere exposure effect* [32], and it has been extensively demonstrated to occur also in situations of interaction with other humans (see ref. [33] for a review). Indeed, the more frequently individuals are exposed to someone, the more they would

be prone to like them and show more willingness to interact with them. Further studies supported it, highlighting that repeated exposure leads to increase perceived likeness, reduces prejudices toward others, and enhances the probability to treat them as social partners, as they are considered as a part of one's own in-group [34–36]. As an explanation, it has been proposed that repeated exposure increases liking because it reduces, over time, people's apprehension toward novelty, such as other humans [33,37]. In other words, humans have evolved to be wary of novel stimuli, which could constitute a potential danger. Therefore, with repeated exposure, individuals gain more knowledge. As they gain more knowledge, they understand that these entities are not inherently threatening. Consequently, over time individuals start to like them more [32,38]. Notably, the same mechanisms seem to take place when interacting with robots, as people report to like robots more and to be well disposed toward them after repeated interactions [23,39].

Alternatively, it may also be that people's affective reaction toward novel entities becomes weaker with their increased familiarity with them, due to the affective habituation [40]. However, it would apply only to "extreme" entities, *i.e.*, entities showing a nearly perfect human representation in terms of physical appearance. According to the *uncanny valley* hypothesis [41], these entities, if still distinguishable from real humans, could amplify people's emotional response toward them. However, for initially neutral stimuli, increased exposure could make them affectively more positive because of the *mere exposure effect* [32].

Taking into account the role that exposure plays for social acceptance and inclusion, it is crucial to address the role that prior knowledge and technical background have in perception of robots as social partners (and hence social inclusion).

2 Aims

The present study aimed at investigating whether the tendency to perceive robots as social partners is modulated by participants' prior knowledge about technology, in terms of theoretical background and technical education.

Another factor that we examined, as having the potential to modulate the readiness to include robots as social partners, was the way the robots are presented to observers, namely whether they are presented as intentional agents or mere mechanical devices. Malle and colleagues (2001) argued that attribution of intentionality helps people to explain their own and others' behavior

in terms of underlying mental causes [42,43]. Humans are good at detecting intentions: a substantial agreement in judgments emerges when people are asked to differentiate between intentional and unintentional behaviors [44]. For example, we can make accurate judgments of intentional behavior from the simple form of appearance of an agent [45]. It can also happen by observing motor signals [46], structured in goal-directed, intentional actions. According to Searle (1999), the competence in predicting and explaining (human) behavior involves the ability to recognize others as *intentional beings*, and to interpret other minds as having "intentional states" such as beliefs and desires [47]. This is what Dennett refers to as the *Intentional Stance*, *i.e.*, the ascription of intentions and intentional states to other agents in a social context [48,49]. In the context of HRI, there are several studies showing that people treat robots *as if* they were living entities endowed with mental states, such as intentions, beliefs, and desires (*e.g.*, [50–52]) following Searle's definition. Interestingly, also form and appearance can relate to the perception of intentionality. For instance, when interacting with an anthropomorphic robot, the likelihood of building a model of its mind increases with its perceived human-likeness [50]. Moreover, people empathize more strongly with human-like robots, so that when human-likeness increases, people's adoption of Intentional Stance toward a robot could be very similar to the one toward a human [53].

A possible explanation of why people might adopt the *Intentional Stance* toward robots is that people are not well informed about how the system has been designed to behave. Thus, they would treat robots as intentional systems, as it allows for using the familiar and well-trained "schema" – usually used to explain other humans' behavior – to explain also the robot's behavior [54,55]. In line with this, it might be that the more people are exposed to robots, the more they would gain knowledge about how these systems are designed and controlled [23]. Therefore, it might prevent people from adopting the *Intentional Stance* and lead them to consider robots only as pre-programmed mechanical systems, making people less willing to perceive them as social partners.

However, to the best of our knowledge, no studies previously investigated how social inclusion depends on the combined effect of both prior knowledge about technology and attribution of intentionality elicited by how the robot is presented to observers.

In this study, addressing this question, we orthogonally manipulated both factors. To test the effect of prior knowledge about technology, we recruited two groups of participants varying in their level of prior knowledge.

Namely, we tested a group of participants with prior knowledge about technology, in terms of theoretical background and technical education (*i.e.*, "Technology Expert" group), and a group of participants having little prior knowledge regarding technology, given their formal education (*i.e.*, "General Population" group).

To evoke different degrees of attribution of intentionality, as a between-subject manipulation we presented participants with a video depicting the iCub robot performing either a goal-directed (intentional) action ("Mentalistic" video; see Data Availability section to access the URL of the video, filename: "Ment_Video.mp4") or a video of a robot being mounted on its platform and then calibrated ("Mechanistic" video, see Data Availability section to access the URL of the video, filename: "Mech_Video.mp4").

To test individuals' tendency to include the robot as an in-group social partner we developed a modified version of the Cyberball task (*e.g.*, [18,19,23]), a well-established task measuring (implicitly) social inclusion (see also [56], for more information). In the original study [19], participants were told that the Cyberball task was simply used to assess their mental visualization skills. The authors found that although participants played with animated icons depicted on the screen, and not with real people, they cared about the extent to which they were included in the game by the other players. For example, if participants were included (*i.e.*, if they received the ball for one-third of the tosses), after the game they reported more positive feelings – in terms of control, self-esteem, and meaningful existence – than if they received the ball only for one-sixth of the tosses [19].

In our version of the Cyberball task, participants were instructed to toss the ball as fast as possible to one of the other players (either iCub or the other human player), being free to choose which player they wanted to toss the ball to. Notably, both players (*i.e.*, the iCub robot and the other human player) were avatars that participants believed to be real agents playing online with them. In detail, the avatar of the iCub robot was programmed to equally alternate the ball between the two players, whereas the avatar of the other human player was programmed to toss the ball to iCub only twice at the beginning of the game, and not at all thereafter. It was intended to make participants believe that the robot was excluded by the other agent, and thus to investigate whether participants tended to interact with iCub by re-including it in the game.

To test the effect of presentation of the robot as either intentional or mechanistic, we asked participants to perform the Cyberball task in two separate sessions, namely before and after watching the "Mechanistic" or "Mentalistic" video

(*i.e.*, Cyberball Pre *vs* Post). Notably, the structure of the Cyberball task was identical in the two sessions.

We hypothesized that if prior knowledge about technology is the sole factor that affects the social inclusion of robots, then people with prior knowledge about technology (*i.e.*, "Technology Expert" sample) should socially include the robot more, compared to non-expert participants (*i.e.*, "General Population" sample), regardless of the way the robot was presented in the video. In line with the *mere exposure effect* [32], it would be because a higher degree of prior knowledge about technology may increase the knowledge about technical systems such as robots, and, as a consequence, the perceived likeness of robotic agents.

In contrast, if the attribution of intentionality is the sole factor affecting the social inclusion of robots, then the probability to re-include the robot should be higher for people who observed iCub being presented as an intentional system compared to people who observed iCub being presented as an artificial, pre-programmed artifact, regardless of participants' prior knowledge about technology. Namely, participants should toss the ball to the iCub more frequently following the video of iCub performing goal-directed actions, as relative to the video in which iCub is shown as being mounted on its platform and then calibrated. This effect should be similar for both technology expert and non-expert participants.

Finally, if both factors (*i.e.*, prior knowledge and the way the robot is presented to participants) play a role in the social inclusion of robots, then we would expect to find an interaction between the two. Namely, the way the robot was presented in the videos, *i.e.*, as an artificial or as an intentional system, should modulate the probability to re-include the robot in the game in the Cyberball Post session compared to Pre, but dependent on prior knowledge about technology (*i.e.*, "General Population" *vs* "Technology Expert" sample).

3 Materials and methods

3.1 Participants

One hundred sixty participants were recruited to participate in the study, via the online platform Prolific (https://prolific.co/). Participants were selected based on the following criteria: age range (18–45 years); fluent level of English, to ensure that participants could understand the instructions; handedness (right-handed); and prior knowledge about technology, in terms of theoretical background

and technical education. Specifically, half of the participants ("Technology Expert" sample) were selected based on "Engineering" and "Computer Science" as educational backgrounds, whereas for the other half ("General Population" sample) we excluded these two backgrounds to prevent collecting data from participants already having prior knowledge about technology, given their formal education. To double-check whether the educational background declared by participants corresponded to the one selected via Prolific, before the experiment we explicitly invited participants to indicate their educational background, and whether it was related to robotics. Notably, four participants who fell into the "General Population" sample declared to have a background in robotics when explicitly asked. Therefore, after checking that they had a background in robotics, they were further included in the "Technology Expert" sample.

The study was approved by the Local Ethical Committee (Comitato Etico Regione Liguria) and conducted in accordance with the ethical standards of the World Medical Association (Declaration of Helsinki, 2013). All participants gave informed consent by ticking the respective box in the online form, and they were naïve to the purpose of the experiment. They all received an honorarium of £4.40 for their participation.

3.2 Procedure

The experiment was a 2 (Session: Cyberball Pre *vs* Post, within-subjects) × 2 (Type of Video: Mechanistic *vs* Mentalistic, between-subjects) × 2 (Group: General Population *vs* Technology Experts, between-subjects) design.

At the beginning of the experiment, participants were asked to perform a modified version of the Cyberball (*e.g.*, [18,19,23]), where participants believed to play online with another human player and the humanoid robot iCub (Figure 1).

Each trial started with the presentation of both the human player and the iCub robot, on the right and the left side of the screen, respectively; the participants' name ("You") was displayed at the bottom. The act of tossing the ball was represented by a 1-s animation of a ball. As previously mentioned, iCub was programmed to alternate between the participant and the human avatar, with an equal probability to pass the ball to either of them; conversely, the human player was programmed to toss the ball to iCub only twice at the beginning of the game, and not thereafter. When participants received the ball, before tossing it, they were instructed to wait until their name (*i.e.*, "You") turned from black to red. Then, they had 500 ms to decide which player to toss the ball to. They were asked to be as fast as possible, being free to choose either of the players. To choose the player on the right side ("Human"), participants had to press the "M" key, whereas they had to press the "Z" key to choose the player on the left side ("iCub"). To make sure that participants were not biased by the different locations of the keys, we asked participants to use a standard QWERTY keyboard to perform the task. If participants took more than 500 ms to choose a player, a red "TIMEOUT" statement was displayed on the screen, and the trial was rejected. The task comprised 100 trials in which participants received the ball in both Pre and Post Cyberball sessions. Namely, in both sessions participants had to choose to toss the ball to either of the two players 100 times. Stimuli presentation and response collection were programmed with PsychoPy v2020.1.3 [57].

After performing the Pre-session Cyberball, participants were asked to watch a 40-s video. As a between-subjects manipulation, half of the users watched a video in which iCub was presented as an artificial system, and
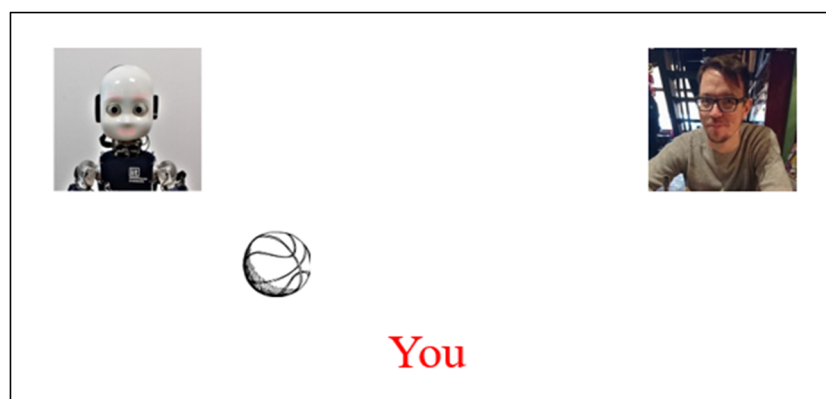

Figure 1: Schematic representation of the Cyberball ball-tossing game.

the other half of users watched a video in which iCub behaved as an intentional, human-like agent performing goal-directed actions.

After watching the videos, participants were asked to answer a few questions (*i.e.*, "How many humans/robots did you see in the video?," and "What title would you give to the video?"). The purpose of these questions was to ensure that participants paid attention to the content of the video. After the experiment, we carefully checked participants' responses to see whether there were responses not congruent with the content of the video. All participants' responses were congruent with the content of the depicted video, indicating that participants paid attention to the video.

After answering the questions, participants were asked to perform the Cyberball again, which was identical to the one performed before the video.

4 Results

4.1 Data preprocessing

Data of two participants, *i.e.*, one from the "General Population" sample and one from the "Technology Expert" sample, were not saved due to a technical error, and therefore they were not included in the analyses. The remaining data were analyzed with R Studio v.4.0.2 [58], using the *lme4* package [59], and JASP Software v.0.14.1 (2020). Data of participants with less than 80% of valid trials (*i.e.*, trials where they pressed either the "Z" or "M" key within 500 ms after participants' "You" name turned red) were excluded from further analyses (11 participants excluded; 5.29% of the total number of trials, mean = 120 ms, SD = 100 ms). Thus, the final sample size on which we ran the analysis was $N = 147$ ("General Population" group, $N = 75$: Mechanistic video, $N = 39$, Mentalistic video, $N = 36$; "Technology Expert" group, $N = 72$: Mechanistic video, $N = 34$, Mentalistic video, $N = 38$). Furthermore, to check for outliers, all trials which deviated ± 2.5 SD from participants' mean Reaction Times (RTs) were excluded from the subsequent analyses (3.17% of trials, mean = 420.86 ms, SD = 26.57 ms).

4.2 Probability of robot choice

To test whether individuals' tendency to include the robot as an in-group social partner was modulated by the combined effect of (i) prior knowledge about technology and (ii) the way the robot was presented, the probability of passing the ball to iCub was considered as the dependent variable in a logistic regression model. Session (Cyberball Pre *vs* Post), Type of Video (Mechanistic *vs* Mentalistic video), and Group (General Population *vs* Technology Experts), plus their interactions, were considered as fixed effects, and Participants as a random effect (see Table 1 for more information about mean values and SD related to the rate of robot choice). Notably, in this study, the model met the assumptions of logistic regression, namely linearity, absence of multicollinearity among predictors, and the lack of strongly influential outliers (see Supplementary Material file, point SM.1, for more information).

Results showed a main effect of Session ($\beta = 0.18$, SE = 0.05, $z = 3.96$, $p < 0.001$, 95% CI = [0.10; 0.28]), with a higher probability to choose the robot in the Cyberball Post-Session compared to Pre-Session. Moreover, a significant Session × Type of Video interaction emerged ($\beta = -0.18$, SE = 0.07, $z = -2.69$, $p = 0.007$, 95% CI = [−0.32; −0.05]).

To investigate this two-way interaction, we first ran two logistic models, with Type of Video (Mechanistic *vs* Mentalistic video) as a fixed effect and Participants as a random effect, separately, according to the within-subject Session (Cyberball Pre *vs* Post). Results showed that, in the Post Session, participants tended to re-include iCub more in the task only after watching the Mechanistic video ($\beta = -0.13$, SE = 0.04, $z = -3.4$, $p < 0.001$, 95% CI = [−0.21; −0.06]; mean values = 58% *vs* 55.7% for Mechanistic and Mentalistic Video, respectively). Importantly, this was not observed in the Cyberball Pre Session ($\beta = 0.10$, SE = 0.07, $z = 0.26$, $p = 0.79$, 95% CI = [−0.06; 0.09]; mean values = 52.2% *vs* 51.8% for Mechanistic and Mentalistic Video, respectively) (Figure 2).

Moreover, we hypothesized that people who observed iCub being presented as an intentional, human-like system (*i.e.*, "Mentalistic" video), compared to people who observed iCub being presented as an artificial, pre-programmed artifact (*i.e.*, "Mechanistic" video) should tend to re-include iCub

Table 1: Mean values and SDs (in parentheses) related to the rate of robot choice, reported separately by Session (Cyberball Pre *vs* Post), Type of Video (Mechanistic *vs* Mentalistic Video), and Group (General Population *vs* Technology Experts)

| Rate of robot choice		Pre	Post
General population	Mechanistic	51.5% (12.2)	56.6% (17.8)
	Mentalistic	50.6% (16.2)	53.7% (17)
Technology experts	Mechanistic	53% (8.5)	59.3% (16.5)
	Mentalistic	52.9% (10.3)	57.7% (14.9)
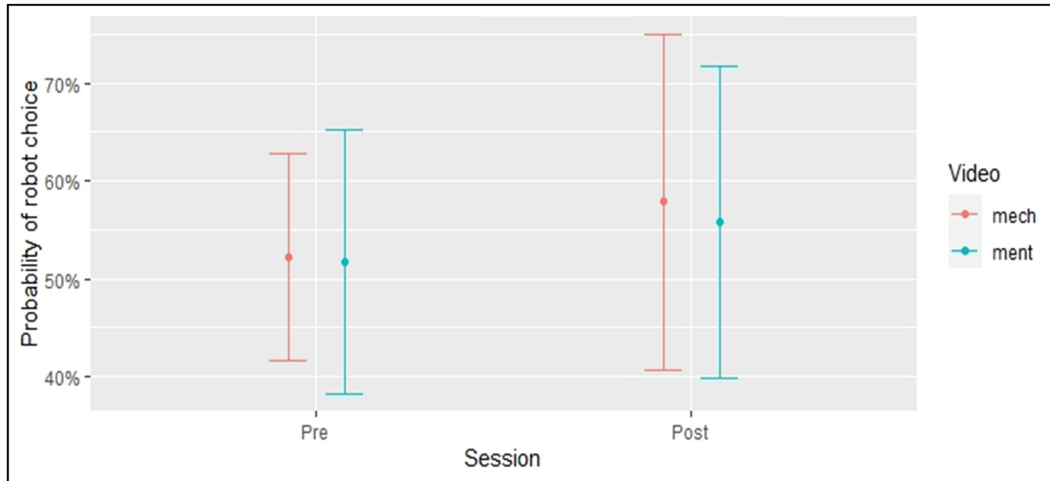
Figure 2: Probability of robot choice as a function of Type of Video (Mechanistic *vs* Mentalistic), plotted separately according to the Session (Pre, on the left side; Post, on the right side).

more in the game. Thus, we also ran two logistic models separately according to Type of Video (Mechanistic *vs* Mentalistic video), with Session (Pre *vs* Post) as a fixed effect and Participants as a random effect. Results showed that, in the Post session compared to Pre, participants tended to re-include iCub more in the game after watching the Mechanistic video (β = 0.20, SE = 0.04, z = 5.77, p < 0.001, 95% CI = [0.13; 0.27], mean values = 52.2% *vs* 58% for Pre and Post sessions, respectively). However, this was not observed for participants who watched the Mentalistic video (β = −0.03, SE = 0.07, z = 1.9, p = 0.06, 95% CI = [−0.02; 0.13]; mean values = 51.8 % *vs* 55.7% for Pre and Post sessions, respectively) (Figure 3).

Notably, the two-way Group × Type of Video interaction resulted not to be significant (β = −0.02, SE = 0.06, z = 0.4, p = 0.68, 95% CI = [−0.1; 0.16]; Table 1), showing that the degree of participants' prior knowledge about technology did not influence the probability of robot choice according to the Type of Video (Mechanistic *vs* Mentalistic video).

Moreover, also the two-way Group × Session interaction was not significant (β = −0.02, SE = 0.07, z = 0.37, p = 0.71, 95% CI = [−0.1; 0.16]; Table 1), showing that participants' degree of prior knowledge about technology did not modulate the probability of robot choice across sessions (Cyberball Pre *vs* Post).
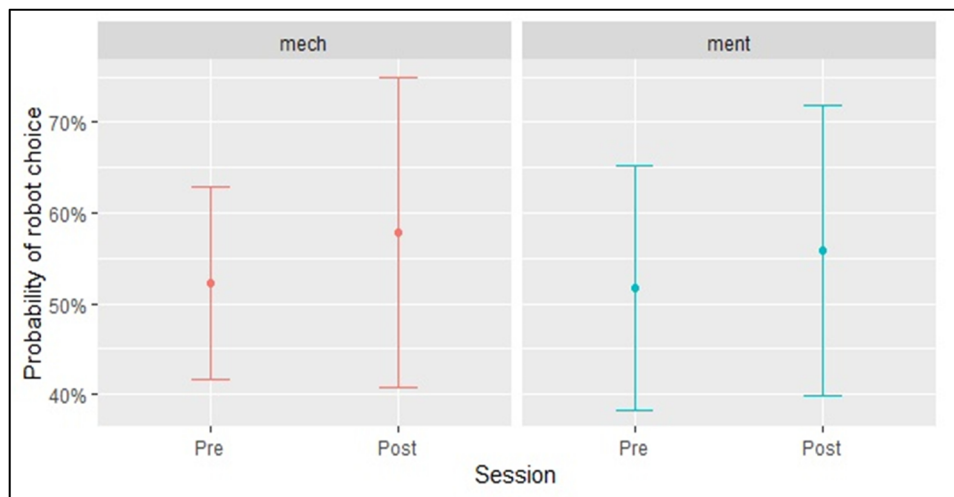
Figure 3: Probability of robot choice as a function of Session (Pre *vs* Post), plotted separately according to the Type of Video (Mechanistic Video, on the left panel; Mentalistic Video, on the right panel).

No other main effect or interaction reached the significance level, with all p-values > 0.31.

5 Discussion

The present study aimed at investigating whether individuals' tendency to include the robot as an in-group social partner would be modulated by (1) the degree of prior knowledge about technology, given participants' theoretical background and technical education, and (2) the way iCub was presented to observers (as an artificial, pre-programmed system vs an intentional agent). Concerning the first aim, we collected two samples of participants varying in their degree of prior knowledge about technology, namely a sample of "Technology Experts" and a "General Population" sample with little prior knowledge about technology. To address our second aim, we asked participants to either watch a "Mechanistic" video, in which iCub was represented as a mechanical artifact, or a "Mentalistic" video, in which iCub was performing a goal-directed action. The tendency to socially include iCub as an in-group member was operationalized as the probability to toss the ball toward the robot during the Cyberball task (e.g., [18,19,23]). Specifically, we asked participants to perform the task in two separate sessions, $i.e.$, before and after watching the videos, to assess whether the behavior of the robot displayed in the video modulated the tendency to re-include iCub in the task.

Our results showed that participants tended to re-include the robot more in the Cyberball Post Session compared to the Pre Session, but only after watching iCub depicted as an artificial system ("Mechanistic" video). It was not the case of people watching iCub presented as an intentional agent ("Mentalistic" video), as they did not show any difference in the probability of robot choice across sessions (Cyberball Pre vs Post). Notably, these effects were not modulated by the degree of individuals' prior knowledge about technology, as the three-way interaction (Session × Type of Video × Group) was not significant.

Therefore, our results suggested that the way the robot was presented to observers, but not the prior knowledge about technology, modulated individuals' tendency to perceive the robot as a social partner. This was also confirmed by the fact that the probability of re-including iCub in the game varied only in the Cyberball Post Session, namely after participants watched iCub in the video.

The results showing that participants were more prone to socially include the robot in the game only after watching the "Mechanistic" video does not support our initial hypothesis. Indeed, we expected that people who observed iCub being presented as an intentional, human-like system ($i.e.$, "Mentalistic" video) should re-include the robot more in the game compared to people who observed iCub being presented as an artificial, pre-programmed artifact ($i.e.$, "Mechanistic" video).

One possible explanation might be that people's attitudes toward robots are driven by their preconceived expectations toward them, much like when interacting with other humans [60]. For example, Marchesi and colleagues [61] recently investigated whether the type of behavior displayed by the humanoid iCub robot affected participants' tendency to attribute mentalistic explanations to the robot's behavior. Thus, they assessed the ascription of intentionality toward robots both before and after participants' observation of two types of behavior displayed by the robot ($i.e.$, decisive vs hesitant). Interestingly, they found that higher expectations toward robots' capabilities might lead to a higher intentionality attribution, with increased use of mentalistic descriptions to explain the robot behavior, even if it was presented as mechanistic [61].

This reasoning is also in line with previous findings in HRI [62,63], which suggest that when people experience unexpected behaviors displayed by robots, the positive or negative value of the violation of expectations may significantly affect individuals' perception of robots as social partners.

In the present study, a possible explanation might be that if people conceive robots as artificial systems, seeing the robot presented in this way ($i.e.$, "Mechanistic" video) might confirm their existing expectations toward robots. Therefore, it could be that if people watched a video in which iCub behaved in the way they expected it to behave, they would be more prone to interact with iCub during the Cyberball game.

An alternative explanation may derive from the $Computer$ Are $Social$ $Actors$ (CASA) framework [64,65]. Originating from the Media Equation Theory [54], it suggests that humans treat media agents, including social robots, like real people, applying scripts for interacting with humans to the interactions with technologies [66]. Importantly, CASA does not apply to every machine or technology: two essential criteria must be respected for a technology serving CASA application [65]. The first criterion is social cues, namely, individuals must be presented with an object that has enough cues to lead the person to categorize it as worthy of social responses [65]. The second criterion is sourcing. Nass and Steuer (1993) clarified that CASA tests whether individuals can be induced to make

attributions toward computers as if they were autonomous sources [67]; namely, whether they can be perceived as an active source of communication, rather than merely transmitting it or only serving as a channel for human–human communication (*e.g.*, [68]). In the light of this, it might be that participants of this study displayed more willingness to interact with the robot only after watching the Mechanistic video because they perceived it as behaving autonomously, thus respecting the second criterion. Related to the first criterion (*i.e.*, presence of social cues), it is important to point out that the perception of what is social varies from person to person and from situation to situation [26]. Therefore, it is difficult to clearly define an objective, universal set of parameters of what constitutes "enough" for signals to be treated as social.

Notably, the CASA framework has argued in favor of the potential role of individual differences such as education or prior experience with technology. Nass and Steuer first argued that some individual differences, including demographics (*e.g.*, level of education) and knowledge about technology might be crucial when testing CASA [67]. In line with this, also recent findings suggest that CASA effects are moderated by factors such as previous computer experience [69], and that people's expectations of *media agents* such as social robots might vary based on their experience [70]. Thus, prior experience with technology seems to be relevant to CASA's assumptions. However, our results are not entirely in line with the predictions stemming from this framework, as our results showed no effect of prior knowledge about technology on participants' tendency to socially include the robot in the Cyberball task. However, these results need to be further confirmed by future studies, also conducted in a well-controlled laboratory setting. In addition, post-experiment questionnaires might be added at the end of the experiment, to disentangle the specific roles of each factor for the social inclusion of robots.

6 Conclusions

Taken together, these findings suggest that the way the robot was presented to observers, but not the degree of prior knowledge about technology, modulated individual tendency to include the robot as an in-group social partner. However, these first exploratory findings need to be addressed in future studies, in more controlled laboratory experiments (as opposed to online testing protocols).

Author contributions: C.R. designed the study, collected and analyzed the data, discussed and interpreted the results, and wrote the manuscript. F.C. designed the study, discussed and interpreted the results, and wrote the manuscript. A.W. designed the study, discussed and interpreted the results, and wrote the manuscript. All the authors revised the manuscript.

References

[1] H. Tajfel and J. C. Turner, "An integrative theory of intergroup conflict," In: W. G. Austin, S. Worchel, editors. *The Social Psychology of Intergroup Relations*. Pacific Grove, CA, Brooks/Col, 1979.

[2] K. Hugenbert and D. F. Sacco, "Social categorization and stereotyping: How social categorization biases person perception and face memory," *Soc. Personal. Psychol. Compass*, vol. 2, no. 2, pp. 1052–1072, 2008.

[3] A. G. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, no. 2, pp. 81–97, 1956.

[4] A. E. Stahl and L. Feigenson, "Social knowledge facilitates chunking in infancy," *Child. Dev.*, vol. 85, no. 4, pp. 1477–1490, 2014.

[5] J. W. Sherman, C. N. Macrae, and G. V. Bodenhausen, "Attention and stereotyping: Cognitive constraints on the construction of meaningful social impression," *Eur. Rev. Soc. Psychol.*, vol. 11, no. 1, pp. 145–175, 2000.

[6] D. E. Broadbent, "The magic number seven after fifteen years," In: A. Kennedy, A. Wilkes, editors. *Studies in Long-term Memory*. London, Wiley, 1975, pp. 3–18.

[7] Van Twuyver and A. Van Knippenberg, "Social categorization as a function of relative group size," *Br. J. Soc. Psychol.*, vol. 38, no. 2, pp. 135–156, 1999.

[8] S. T. Fiske and S. L. Neuberg, "A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation," *Adv. Exp. Soc. Psychol.*, vol. 23, pp. 1–74, 1990.

[9] D. P. Skorich, K. I. Mavor, S. A. Haslam, and J. L. Larwood, "Assessing the speed and ease of extracting group and person information from faces," *J. Theor. Soc. Psychol.*, vol. 5, pp. 603–23, 2021.

[10] J. Krueger, "The psychology of social categorization," In: N. J. Smelser and P. B. Baltes, editors. *The international encyclopedia of the social and behavioral sciences*. Amsterdam, Elsevier; 2001.

[11] C. N. Macrae and G. V. Bodenhausen, "Social cognition: thinking categorically about others," *Annu. Rev. Psychol.*, vol. 51, no. 1, pp. 93–1, 2000.

[12] L. Castelli, S. Tomelleri, and C. Zogmaister, "Implicit ingroup metafavoritism: Subtle preference for ingroup members displaying ingroup bias," *Per Soc. Psychol. Bull.*, vol. 34, no. 6, pp. 807–818, 2008.

[13] D. M. Buss, "Do women have evolved mate preferences for men with resources? A reply to Smuts," *Ethol. Sociobiol.*, vol. 2, no. 5, pp. 401–408, 1991.

[14] L. A. Duncan, J. H. Park, J. Faulkner, M. Schallen, S. L. Neuberg, and D. T. Kenrick, "Adaptive allocation of attention: effects of sex and sociosexuality on visual attention to attractive opposite-sex faces," *Evol. Hum. Behav.*, vol. 28, no. 5, pp. 359–364, 2007.

[15] R. Cordier, B. Milbourn, R. Martin, A. Buchanan, D. Chung, and D. Speyer, "A systematic review evaluating the psychometric properties of measures of social inclusion," *PLoS One*, vol. 12, no. 6. p. e0179109, 2017.

[16] A. Wykowska, "Social robots to test flexibility of human social cognition," *Int. J. Soc. Robot.*, vol. 12, no. 6, pp. 1203–1211, 2020.

[17] F. Eyssel and F. Kuchenbrandt, "Social categorization of social robots: anthropomorphism as a function of robot group membership," *Br. J. Soc. Psychol.*, vol. 51, no. 4, pp. 724–731, 2012.

[18] K. D. Williams, C. C. K. Cheung, and W. Choi, "Cyberostracism: effects of being ignored over the internet," *J. Pers. Soc. Psychol.*, vol. 9, no. 5, pp. 748–762, 2000.

[19] K. D. Williams and B. Jarvis, "Cyberball. A program for use in research on interpersonal ostracism and acceptance," *Behav. Res. Methods*, vol. 38, no. 1, pp. 174–180, 2006.

[20] F. Bossi, M. Gallucci, and P. Ricciardelli, "How social exclusion modulates social information processing: a behavioural dissociation between facial expressions and gaze direction," *PLoS One*, vol. 13, no. 4, p. e0195100, 2018.

[21] I. Van Beest and K. D. Williams, "When inclusion costs and ostracism pays, ostracism still hurts," *J. Pers. Soc. Psychol.*, vol. 91, no. 5, pp. 918–928, 2006.

[22] A. R. Cartell-Sowell, Z. Chen, and K. D. Williams, "Ostracism increases social susceptibility," *Soc. Influ.*, vol. 3, no. 3, pp. 143–153, 2008.

[23] F. Ciardo, D. Ghiglino, C. Roselli, and A. Wykowska, "The effect of individual differences and repetitive interactions on explicit and implicit measures towards robots," In: A. R. Wagner, et al. editors. *Social robotics. ICSR 2020: Lecture Notes in Computer Science; 2020 Nov 14–18.; Golden, Colorado*. Cham: Springer, 2020, pp. 466–477.

[24] M. J. Bernstein, D. F. Sacco, S. G. Young, K. Hugenberg, and E. Cook, "Being "in" with the in-crowd: The effects of social exclusion and inclusion are enhanced by the perceived essentialism of ingroups and outgroups," *Pers. Soc. Psychol. Bull.*, vol. 36, no. 8, pp. 999–1009, 2010.

[25] A. B. Allen and W. K. Campbell, *Individual Differences in Responses to Social Exclusion: Self-esteem, Narcissism, and Self-compassion*. In: N. C. DeWall, editor. UK, Oxford University Press; 2013, pp. 220–227.

[26] A. Waytz, J. Cacioppo, and N. Epley, "Who sees human? The stability and importance of individual differences in anthropomorphism," *Perspect. Psychol. Sci.*, vol. 5, no. 3, pp. 219–232, 2010.

[27] N. A. Hinz, F. Ciardo, and A. Wykowska, "Individual differences in attitude toward robots predict behavior in human-robot interaction," M. Salichs, et al., editors. *Social Robotics. ICSR 2019: Lecture Notes in Computer Science; 2019 Nov 26–29*, Madrid, Spain, Cham: Springer; 2019, pp. 64–73.

[28] M. Heerink, "Exploring the influence of age, gender, education and computer experience on robot acceptance by older adults," *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 2011 Mar 6–9*. Lausanne, Switzerland, IEEE; 2011.

[29] D. Li, P. P. L. Rau, and D. Li, "A cross-cultural study: effect of robot appearance and task," *Int. J. Soc. Robot.*, vol. 2, no. 2, pp. 175–186, 2010.

[30] S. Marchesi, C. Roselli, and A. Wykowska, "Cultural values, but not nationality, predict social inclusion of robots," In: H. Li, et al., editors. *Social Robotics. ICSR 2021: Lecture Notes in Computer Science; 2021 Nov 10–13*, Singapore. Cham, Springer. 2021, pp. 48–57.

[31] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, "The iCub humanoid robot: an open platform for research in embodied cognition," *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems; 2008 Aug 19–21; Gaithersburg, Maryland*. New York: Association for Computing Machinery; 2008.

[32] R. B. Zajonc, "Attitudinal effects of mere exposure," *J. Pers. Soc. Psychol.*, vol. 9, no. 2, pt.2, pp. 1–27, 1968.

[33] R. F. Bornstein, "Exposure and affect: overview and meta-analysis of research, 1968–1987," *Psychol. Bull.*, vol. 106, no. 2, pp. 265–289, 1989.

[34] K. Mrkva and L. Van Boven, "Salience theory of mere exposure: relative exposure increases liking, extremity, and emotional intensity," *J. Pers. Soc. Psychol.*, vol. 118, no. 6, pp. 1118–1145, 2020.

[35] L. A. Zebrowitz, B. White, and K. Wieneke, "Mere exposure and racial prejudice: exposure to other-race faces increases liking for strangers of that race," *Soc. Cogn.*, vol. 26, no. 3, pp. 259–275, 2008.

[36] M. Brewer and N. Miller, "Contact and cooperation," In: P. A. Katz and D. A. Taylor, editors. *Eliminating racism. Perspectives in Social Psychology* (*A Series of Texts and Monographs*), Boston, MA, Springer, 1988.

[37] A. A. Harrison, "Mere exposure. In Advances in experimental social psychology," *Adv. Exp. Soc. Psychol.*, vol. 10, pp. 39–83, 1997.

[38] R. M. Montoya, R. S. Horton, J. L. Vevea, M. Citkowicz, and E. A. Lauber, "A re-examination of the mere exposure effect: the influence of repeated exposure on recognition, familiarity, and liking," *Psychol. Bull.*, vol. 143, no. 5, pp. 459–498, 2017.

[39] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura, "The influence of people's culture and prior experiences with Aibo on their attitude towards robots," *AI Soc.*, vol. 21, no. 1–2, pp. 217–230, 2007.

[40] J. A. Zlotowski, H. Sumioka, S. Nishio, D. F. Glas, C. Bartneck, and H. Ishiguro, "Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception," *Front. Psychol.*, vol. 6, p. 883, 2015.

[41] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley," *IEEE Robot. Autom. Mag.*, vol. 19, no. 2, pp. 98–100, 2012.

[42] B. F. Malle, L. J. Moses, and D. A. Baldwin, "The significance of Intentionality," In: B. F. Malle, L. J. Moses, and D. A. Baldwin, editors. *Intentions and Intentionality: Foundations of Social Cognition*, Cambridge, MA, MIT Press, 2001.

[43] S. Thellman, A. Silvervarg, and T. Ziemke, "Folk-psychological interpretation of human vs humanoid robot behavior: exploring the intentional stance toward robots," *Front. Psychol.*, vol. 8, p. 1962, 2017.

[44] B. F. Malle and J. Knobe, "The folk concept of intentionality," *J. Exp. Soc. Psychol.*, vol. 33, no. 2, pp. 101–121, 1997.

[45] D. Morales-Bader, R. D. Castillo, C. Olivares, and F. Miño, "How do object shape: semantic cues, and apparent velocity affect the attribution of intentionality to figures with different types of movements?," *Front. Psychol.*, vol. 11, p. 935, 2020.

[46] H. C. Barrett, P. M. Todd, G. F. Miller, and P. W. Blythe, "Accurate judgments of intention from motion cues alone: a cross-cultural study," *Evol. Hum. Behav.*, vol. 26, pp. 313–331, 2005.

[47] J. R. Searle, *Mind, Language and Society: Philosophy in the Real World*. New York, NY, Basic Books, 1999.

[48] D. C. Dennett, "Intentional systems," *J. Philos.*, vol. 68, no. 4, pp. 87–106, 1971.

[49] D. C. Dennett. *The Intentional Stance*. Cambridge, MA, MIT Press; 1989.

[50] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher, "Can machines think? Interaction and perspective taking with robots investigated via Fmri," *PLoS One*, vol. 3, p. e2597, 2008.

[51] A. Waytz, C. K. Morewedge, N. Epley, G. Monteleone, J. H. Gao, and J. T. Cacioppo, "Making sense by making sentient: effectance motivation increases anthropomorphism," *J. Pers. Soc. Psychol.*, vol. 99, pp. 410–435, 2010.

[52] S. Marchesi, D. Ghiglino, F. Ciardo, J. Perez-Osorio, E. Baykara, and A. Wykowska, "Do we adopt the intentional stance toward humanoid robots? *Front. Psychol.*, vol. 10, p. 450, 2019.

[53] J. Perez-Osorio and A. Wykowska, "Adopting the intentional stance toward natural and artificial agents," *Philos. Psychol.*, vol. 33, pp. 369–395, 2020.

[54] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media like Real People*. Cambridge, UK, Cambridge University Press; 1996.

[55] S. L. Lee, I. Y. M. Lau, S. Kiesler, and C. Y. Chiu, "Human mental models of humanoid robots," *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (*ICRA*); *Apr 18–22*. Barcelona, Spain, IEEE, 2005.

[56] L. Mwilambwe-Tshilobo and R. N. Spreng, "Social exclusion reliably engages the default network: a meta-analysis of Cyberball," *NeuroImage*, vol. 227, p. 117666, 2021.

[57] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, et al., "PsychoPy2: Experiments in behavior made easy," *Behav. Res. Methods*, vol. 51, pp. 195–203, 2019.

[58] Team RC. R: A Language and Environment for Statistical Computing. http://www.R-project.org/.

[59] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. Christensen, et al. "Package 'lme4'. Linear mixed-effects models using S4 classes," *R Package version*, vol. 1, no. 6, 2011 Mar 7.

[60] V. Lim, M. Rooksby, and E. S. Cross, "Social robots on a global stage: establishing a role for culture during human–robot interaction," *Int. J. Soc. Robot.*, vol. 13, no. 6, pp. 1307–1333, 2021.

[61] S. Marchesi, J. Pérez-Osorio, D. De Tommaso, A. Wykowska, "Don't overthink: fast decision making combined with behavior variability perceived as more human-like," *2020 29th IEEE International Conference on Robot and Human Interactive Communication* (*RO-MAN*); *2020 Aug 31- Sep 4*. Naples, Italy, IEEE, 2020.

[62] H. Claure and M. Jung, "Fairness considerations for enhanced team collaboration," *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (*HRI*); *2021 Mar 9–11*. IEEE, 2021.

[63] J. K. Burgoon, "Interpersonal expectations, expectancy violations, and emotional communication," *J. Lang. Soc. Psychol.*, vol. 12, no. 1–2, pp. 30–48, 1993.

[64] C. Nass, J. Steuer, E. R. Tauber, "Computers are social actors," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 1994 Apr 24–28; Boston, Massachusetts*. New York, Association for Computing Machinery; 1994.

[65] C. Nass and Y. Moon, "Machines and mindlessness: social responses to computers," *J. Soc. Issues*, vol. 56, no. 1, pp. 81–103, 2000.

[66] A. Gambino, J. Fox, and R. A. Ratan, "Building a stronger CASA: Extending the computers are social actors paradigm," *Hum. Mach. Commun. J.*, vol. 1, pp. 71–85, 2020.

[67] C. Nass and J. Steuer, "Voices, boxes, and sources of messages: computers and social actors," *Hum. Commun. Res.*, vol. 19, no. 4, pp. 504–527, 1993.

[68] S. S. Sundar and C. Nass, "Source orientation in human-computer interaction: programmer, networker, or independent social actor? *Commun. Res.*, vol. 27, pp. 683–703, 2000.

[69] D. Johnson and J. Gardner, "The media equation and team formation: further evidence for experience as a moderator," *Int. J. Hum. Comput.*, vol. 65, pp. 111–124, 2007.

[70] A. C. Horstmann and N. C. Krämer, "Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment," *Front. Psychol.*, vol. 10, p. 939, 2019.