



Shared Knowledge in Human-Robot Interaction (HRI)

Laura Miraglia¹ · Cinzia Di Dio^{1,2} · Federico Manzi^{1,2} · Takayuki Kanda^{3,4} · Angelo Cangelosi⁵ · Shoji Itakura⁶ · Hiroshi Ishiguro^{4,7} · Davide Massaro^{1,2} · Peter Fonagy⁸ · Antonella Marchetti^{1,2}

Accepted: 19 July 2023
© The Author(s) 2023

Abstract

According to the Theory of Natural Pedagogy, object-directed emotion may provide different information depending on the context: in a communicative context, the information conveys culturally shared knowledge regarding the emotional valence of an object and is generalizable to other individuals, whereas, in a non-communicative context, information is interpreted as a subjective disposition of the person expressing the emotion, i.e., personal preference. We hypothesized that this genericity bias, already present in infants, may be a feature of human communication and, thus, present at all ages. We further questioned the effects of robotic ostensive cues. To explore these possibilities, we presented object-directed emotions in communicative and non-communicative contexts under two conditions: adult participants ($N=193$) were split into those who underwent the human-demonstrator condition and those who underwent the robot-demonstrator condition, i.e., a human actor or a robot displayed object-directed emotions, respectively. Questionnaires further assessed the participants' attachment style and mentalization ability. The results showed that (1) Natural Pedagogy Theory applies to humans across the lifespan; (2) Shared knowledge depends on the contexts (communicative vs. non-communicative) and who is sharing the information (human or robot); and (3) robotic ostensive cues trigger participants' attention, conversely, in their absence, participants do not turn the robot into a communicative partner by not assigning it a communicative intention due to a difficulty in reading the robot's mind. Taken together, our findings indicate that robotic ostensive cues may ease the human-robot interaction (HRI), which is also biased by the human attachment style. The study has been preregistered in Open Science Framework, OSF on September 9, 2021 (Registration DOI <https://doi.org/10.17605/OSF.IO/9TWY8>).

Keywords Human-robot interaction (HRI) · Natural pedagogy theory · Shared knowledge · Epistemic trust · Social cognition · Attachment style

Cinzia Di Dio and Federico Manzi contributed equally to this work.

✉ Laura Miraglia
laura.miraglia@unicatt.it

¹ Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Largo Gemelli 1, Milano 20123, Italy

² Research Unit in Psychology and Robotics in the Lifespan (PsyRoLife), Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

³ Human-Robot Interaction Laboratory, Department of Computer Science, Kyoto University, Kyoto, Japan

⁴ Advanced Telecommunications Research Institute International, IRC/HIL, Keihanna Science City, Kyoto, Japan

⁵ School of Computer Science, The University of Manchester, Manchester, UK

⁶ Centre for Baby Science, Doshisha University, Kyoto, Japan

⁷ Department of Systems Innovation, Osaka University, Toyonaka, Japan

⁸ Research Department of Clinical, Educational and Health Psychology, UCL, London, UK

1 Introduction

1.1 Natural Pedagogy Theory

The evolutionary success of our species depends crucially on the social transmission of knowledge both contemporaneously and throughout historical time. Hence, one of the first challenges a human being faces is learning *about* and *from* the world around him/her. Children are able to draw information relevant to their behavior by simply observing the reactions of adults and gaze direction toward an object or event [1, 2], and by decoding emotional information and discriminating between facial [3] and vocal expressions [2]. This form of social cognition emerges at approximately one year of age when infants begin to engage with others in various types of joint attentional activities, such as gaze, social reference, and gestural communication, which generate cultural learning that enables the acquisition of language, discursive skills, tool-use practices, and other conventional activities [4, 5].

Csibra and Gergely [6, 7] framed out the Natural Pedagogy Theory which posits *Pedagogy* as a specific type of communication that enables rapid and efficient social learning that – similarly to all types of social learning (imitation, emulation, etc.) – conveys generalizable knowledge that is valid beyond the actual situation [6]. Thus, it is important to consider the distinctive nature of *Pedagogy* both as a particular type of social learning and as a particular type of communication. Csibra and Gergely's theory is grounded on the Gricean notion of ostensive communication, which postulates that an essential feature of human communication is the expression and recognition of intents [8]. Ostensive communication is achieved through the production of ostensive signals, stimuli, or cues that indicate a communicative intention towards an addressee. Ostensive cues typically lead the addressee to feel recognized as a subject [9], encouraging more rapid knowledge acquisition [10] and allowing the establishment of epistemic trust [9, 11]. Importantly, it is plausible that secure attachment acts as a guarantee of the authenticity of knowledge [12], as the child is more likely to attend to the known and trusted adult indicating and naming new objects or showing whether the object is good or bad through social referencing [3, 4].

1.2 Ostensive Cues and Social Interaction

Human infants are highly sensitive to social cues [2–4, 13, 14], i.e., behavioral cues – such as eye contact, infant-directed speech, turn-taking contingent discourse, calling an infant by name, etc. – which indicate a clear communicative intention of an agent. This has led to a growing interest in how social cues, such as object-directed emotional

expressions, can be an important source of social information. The caregiver's ostensive cues not only cause the infant to interpret the adult's action as indicative of a communicative intention to transfer relevant knowledge but also generate attachment security through a sensitive and contingent response [12]. Social cues act differently in terms of preparing the observer to obtain certain types of object information and should be distinguished from non-communicative cues concerning the expected effects [6, 7]. In this respect, in communicative contexts, i.e., when the addressee is engaged through ostensive cues, a *genericity bias* is generated, in other words, the information conveyed is processed as generalizable to other individuals and valid beyond the present situation [7, 15–17]. In this sense, ostensive cues make it possible to convey generally shared knowledge. Infants' sensitivity to ostensive signals triggers an automatic predisposition in the child to receive new and relevant information, made manifest by the communicative intention of the adult through ostensive communication [18, 19]. Within these communicative contexts, ostensive cues (such as gaze shifting, head movement, and pointing) generate an expectation of generic content in the addressee: unless the context or other cues specify otherwise, children interpret the information they receive as generic rather than episodic [7, 15, 20]. Consequently, Natural Pedagogy theorists have argued that an ostensive or communicative context generates a genericity bias whereby the addressee expects to be taught something generalizable and focuses his or her attention on the intrinsic characteristics of the object referent [7]. For instance, Yoon and colleagues [17] showed that when 9-month-old infants were introduced to an object in a communicative context, they remembered generic properties of the object (such as its identity); but when the same object was presented in non-communicative contexts, they were more likely to remember its location, that is context-specific properties. In line with these results, Marno and colleagues [16] also found that, in a communicative context, adult participants preferentially encoded the object's identity at the expense of its location, showing that communicative cues modulate attention to and encoding of the properties of an object in adults as well. Moreover, a recent study by Okumura et al. [21] reported that, although both attentional cues (such as a beep) and ostensive cues affected infants' gaze-following, only ostensive cues facilitated their referential object learning. In line with these results, studies that have explored the effects of ostensive cues on infants' tendency to follow others' gaze toward objects, further showed that children were more likely to follow the agent's gaze if it was preceded by ostensive cues [22], even when the agent was a robot [23]. In the experiments by Okumura et al. [23], 12-month-olds watched videos in which a human or a robot looked toward an object. Their aim was to examine

whether robots can influence infants' learning and, given the empirical evidence that has demonstrated the importance of verbalizations in establishing joint attention in infant-adult interactions [24], the authors added ostensive verbal signals while the robot gazed at an object. Results showed that when the robot's gaze was accompanied by ostensive verbal cues, children not only followed the direction of the robot's gaze but also paid preferential attention to the object when the ostensive cue was present. Based on this evidence, Natural Pedagogy theorists have argued that children encode information differently depending on whether it is presented in a communicative context compared with a non-communicative context.

1.3 Shared Knowledge

Within the conceptual framework of Shared Knowledge [7], in communicative contexts, children would assign an object-centered interpretation to individuals' object-directed emotions. Namely, the addressee of ostensive communication would focus on the intrinsic characteristics of the object and this kind of interpretation would allow children to (a) act in an emotionally consistent way not only toward the referent in the here and now, but also in the future situations, and (b) expect other people to share the same emotional disposition and act accordingly toward the same type of referent [15]. Ostensive signals increase the likelihood that the information provided will be generalized to other circumstances or interactions [25]. To exemplify the concept, Egyed et al. [15] provided the example of a snake: the parent who sees their child approaching a snake will show an expression of fear towards it to warn the child of the danger. The adult intends to communicate to the child that the snake is dangerous to approach. By addressing the child with ostensive signals, the child will assign to the referent an *object-centered* interpretation generalizable to future situations and other individuals, i.e., an awareness that snakes are dangerous. On the other hand, when the expression of fear is observed in a non-communicative context, infants would assign a *person-centered* interpretation to the snake that would lead them to not generalize the emotional disposition as applicable to other individuals but would interpret the object-directed emotion as an emotional attitude of that person (e.g., my parent is afraid of snakes) [15]. Previous works have suggested that children, even at very early ages, are able to flexibly assign person- and object-centered interpretations to the display of referential emotions depending on whether they are shown in a communicative or non-communicative context [15, 26]. For instance, Egyed and colleagues [15] found that 18-month-old infants flexibly assign person- and object-centered interpretations according to the context in which the emotion was displayed. The experiment consisted

of an actress displaying positive versus negative emotions toward two objects differing in their shape and color; then, the same or another actress made a request for one of the objects. After being addressed in an ostensive communicative manner, infants were more likely to choose the object with a positive valence in response to the unknown actress' request. On the other hand, when the object-directed emotion was displayed within a non-communicative context (i.e., infants were not directly addressed), infants did not generalize the object-directed emotion when responding to the different actress's object request. The results suggested that 18-month-olds interpret expressions of emotion toward an object communicated in an ostensive way as revealing general valence information about the object that is also relevant to and shared by other people. In other words, infants assigned to the object-directed emotion an *object-centered* interpretation. In contrast, when the same emotion expression is displayed in a non-communicative context, infants' interpretation is *person-centered*, i.e., infants interpret the object-directed emotion as a person-specific attitude or a personal preference (she likes it/she does not like it) and episodic.

1.4 Aim of the Study

The aim of the present study is twofold: (1) to investigate whether the *Shared Knowledge assumption* is a feature of human communication and thus it is found in adulthood; (2) to evaluate whether ostensive cues acted upon by a robotic agent may lead to effects beyond mere attentional arousal and whether the Shared Knowledge assumption persists in human-robot interaction (HRI) as well. These questions were inspired by the increasing use of robotic agents in educational settings, which demands an effort in understanding the mechanisms underlying HRI. Numerous studies have contributed to our understanding of how people interact with robots in educational contexts [27–34]. These studies have shown that the attribution of mental abilities and psychological traits to robots [32] (for a review, see also [35]) and the human-likeness, play a significant role in establishing trust and facilitating human-robot interaction [36–41]. Some fundamental mechanisms of social cognition, such as eye gaze [21, 23, 42] and joint attention [43], have been studied using humanoid robots, but little is actually known about the effect of ostensive cues on the conveyance of relevant information and related generalization processes acted by a robot.

Although there is little work in the state of art investigating the hypotheses of Natural Pedagogy in adults, the promising results found by Marmo and colleagues [16] suggest that communicative and non-communicative contexts do not exclusively exert their effects on infants. Based on these

assumptions, we hypothesized that the Shared Knowledge assumption described above reflects a feature of human communication, and if so, it should persist in adulthood. In addition, we wondered whether the Shared Knowledge assumption is restricted only to human ostensive cues. That is, what happens when a robot ostensibly engages a person through eye contact and greetings? What effects do ostensive cues act by a robot exert on adult participants? Our study investigated these questions by directly comparing conditions in which a human or a robot displayed object-directed emotions in both communicative and non-communicative contexts. To investigate these hypotheses, one-hundred and ninety-three (193) Italian adult participants (age range = 18–61 years) were involved in the study. We have, therefore, developed a paradigm inspired by the work of Egyed, Király, and Gergely [15] to test whether the Shared Knowledge assumption persists into adulthood and whether this phenomenon is activated when a robotic agent acts as a communicative partner. We expected to replicate the results of the original work regarding the persistence of the genericity bias in adulthood. With respect to the robot condition, we hypothesized that robot ostensive communication acted by a robot that goes beyond mere attentional arousal and might influence participants' likelihood of sharing the positively valenced object. We have modified the original paradigm [15] to adapt it to adult participants and online administration. In order to ensure optimal control over the actions and behaviors of both the human and robot participants in this initial study, we opted for a video-based version of the interaction, where several parameters could be properly manipulated and controlled.

Assuming a significant effect of the robot agent on the participant's behavior, it would be conceivable to evaluate the behavior in a more ecological setting. We recorded video clips representing humans and robots acting as the demonstrator and requester (the one performing the object-directed emotion) and the requester (the one making the request to share). We split the sample into those who underwent the robot-demonstrator condition and those who underwent the human-demonstrator condition. The actor (human or robot) displayed two different emotions, one with positive and the other with negative valence, toward two different unfamiliar objects. We involved two social robots, namely QT Robot and NAO, to play the role of demonstrator and requester in the experimental condition. Crucially, during the familiarization phase, the demonstrator displayed emotions in a communicative (the human or the robot ostensibly greeted the participant) or non-communicative context (the human or the robot acted as if alone); in the test phase, participants saw the requester make a request by reaching his/its hand (the requester could be the same, a different person/robot or another agent depending on the demonstrator's agency) and, subsequently, chose which object sharing with the requester. After the Shared Knowledge task, we administered two questionnaires assessing participants' attachment style and mentalization ability, the Attachment Style Questionnaire (ASQ; [44]) and the Reflective Functioning Questionnaire (RFQ; [45]) respectively. Participants' attachment style and mentalization ability were assessed because they are constructs intrinsically linked to ostensive communication. Finally, we administered the Attribution of Mental States questionnaire (AMS-Q; [46]), a tool that assesses the degree of mental anthropomorphism of nonhuman agents.

Table 1 Sample socio-demographic characteristics

Sociodemographic characteristics	
Age, mean \pm SD	27.98 \pm 8.89
Gender	N (%)
Male	97 (50.3%)
Female	96 (49.7%)
Residence	N (%)
Northwest Italy	56 (29.0%)
Northeast Italy	43 (22.3%)
Centre Italy	46 (23.8%)
South Italy	23 (11.9%)
Sicily and Sardinia	25 (13.0%)
Educational level	N (%)
Middle school	2 (1.0%)
High school	96 (49.7%)
Graduate school	86 (44.6%)
Postgraduate school	9 (4.7%)
Employment status	N (%)
Student	96 (49.7%)
Employed	74 (38.3%)
Unemployed	17 (8.8%)
Other	6 (3.1%)

2 Methods

2.1 Participants

One-hundred and ninety-three (193) Italian adult participants (*Mean age* = 27.98 years, *SD* = 8.89, *age-range* = 18–61 years) took part in the study. Inclusion criteria for all participants were age of majority and being a native Italian speaker. See Table 1 for details of the demographic characteristics of the sample. The participants were recruited on Prolific for 5.50\$ per hour. The platform allows participants to be selected based on nationality and other characteristics of interest (e.g., absence of special medical conditions). Participants were informed about the experimental procedure, the measurement items, and the materials. All participants gave written informed consent in line with the Declaration of Helsinki and its revisions and in accordance with the requirements of the ethics committee of the Department of

Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, which approved this study.

2.2 Procedure and Task

2.2.1 General Procedure

In the current study, we modified the paradigm used by Egyed and colleagues [15] to adapt it to adult participants and online administration. Participants were assessed under two experimental conditions: the sample was split into those who underwent the robot-demonstrator condition and those who underwent the human-demonstrator condition. In both the human and robot-demonstrator conditions, the experiment began with a familiarization phase in which object-directed emotion displays were presented in a communicative context, i.e., the demonstrator ostensibly engaged with participants through eye contact and greetings, or non-communicative context, i.e., the demonstrator acted as if alone, without looking into the camera nor greeting participants. This was followed by a test phase in which the requester made a request by extending his/its arm toward participants and asking them to give an object. The object that is positively valenced by the demonstrator is referred to as the target object. As in the original work [15], we varied the identity of the requester, who might be the same person who showed the expressions of referential emotions or a different person; in addition, we also varied the genus of the requester which could be a human or a robot. We involved two social robots, i.e., QT Robot and NAO (Fig. 1), to play the role of demonstrator and requester in the experimental condition.

We, therefore, created six experimental conditions: (1) communicative-context/same-person condition; (2) communicative-context/different-person condition; (3) communicative-context/different genus condition; (4) noncommunicative-context/same-person condition; (5)

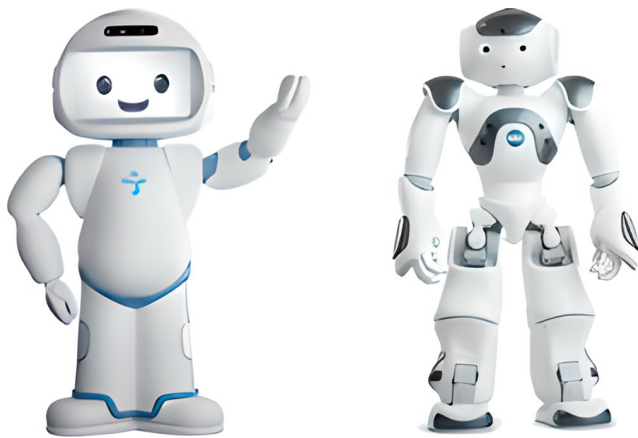


Fig. 1 QT Robot (on the left) and Nao (on the right)

noncommunicative- context/different-person condition; and (6) noncommunicative-context/different genus condition. All conditions were administered in random order and were semi-balanced by agent (human-human, human-robot) and by role (demonstrator, requester). All conditions are illustrated in Figs. 2 and 3.

After the Shared Knowledge task, participants were administered the following tests: the Attribution of Mental States questionnaire (AMS-Q) [46], the Reflective Functioning Questionnaire (RFQ) [45] (Italian version: [47]), and a short version of the Attachment Style Questionnaire (ASQ) [44] (Italian version: [48]). The questionnaires were administered in random order.

2.2.2 Experimental Conditions: Shared Knowledge task

The design of the Shared Knowledge task was a $2 \times 3 \times 2$ repeated measures mixed model, with 2 levels of *context* (communicative, non-communicative), 3 levels of *requester* (*same* identity – i.e. demonstrator and requester were the same person/robot; *different* identity – within agency, i.e., if the demonstrator was a human the requester was another human; if the demonstrator was a robot the requester was another robot; *other* identity – between agency, i.e., if the demonstrator was human, the requester was a robot and vice-versa) as the within-subject factors, and 2 levels of *demonstrator* (human, robot) as the between-subject factor.

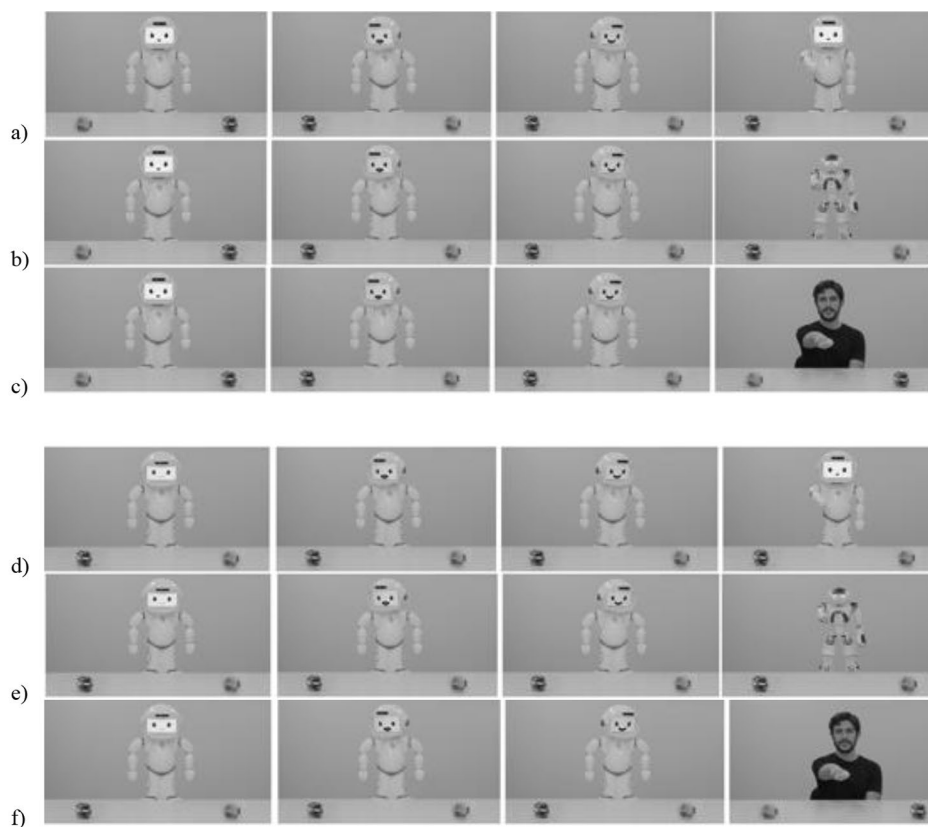
The sample was initially split into two groups (between-subject factor): those who underwent the robot-demonstrator condition and those who underwent the human-demonstrator condition. Within each group, the participant watched six short video clips with a 24-second duration each (each frame of the video has the same duration), showing different conditions. Each condition differed in the type of *context* (communicative vs. non-communicative, i.e., the demonstrator gazing toward and verbally engaging the participant prior to emotions display vs. a non-engaging approach); the demonstrator's *agency* (human vs. robot); and *requester*: same agent (i.e., demonstrator and requester were the same person/robot), different identity (demonstrator and requester had the same agency – human or robot – but a different identity), and finally, other agent (the demonstrator and the requester could be human and robot respectively, or robot and human, thus counterbalanced by role). The experimental condition was as follows:

- *Familiarization phase*: in this initial phase, a human or robotic demonstrator displayed an object-directed emotion, expressing joy toward one object and then turning toward the other, presenting an emotional expression of disgust (the objects are described in *Stimuli*). This sequence was repeated a second time. Before showing

Fig. 2 Human-demonstrator conditions: **(a)** communicative context, same person; **(b)** communicative context, different person; **(c)** communicative context, robot; **(d)** non-communicative context, same person; **(e)** non-communicative context, different person; **(f)** non-communicative context, robot



Fig. 3 Robot-demonstrator conditions: **(a)** communicative context, same robot; **(b)** communicative context, different robot; **(c)** communicative context, human; **(d)** non-communicative context, same robot; **(e)** non-communicative context, different robot; **(f)** non-communicative context, human



the emotions, in the communicative context, the demonstrator ostensibly addressed the participants through eye contact (looking into the camera) and smiling while greeting them, saying, “Hi! Pay attention”. In the non-communicative context, the demonstrator never interacted with the participants: the human or robotic actor never looked at or talked to participants either before or during her object-directed expressions of emotions.

- **Test phase:** In this subsequent phase, the human or robotic requester communicatively addressed the participants using ostensive signals (looking, smiling, greeting); he/it then displayed a hand request gesture (reaching out and placing his/its hand between the two objects with the palm facing upward), and said, “Give me one of them!”. Throughout the test phase, the requester would only look at the camera and never at the objects. At this point, participants had to select which object they would like to share with the requester.

Stimuli. Two unfamiliar objects with different colors, different shapes, and similar affordance properties for both humans and robots were used (about 6.18 inch). Their left-right position on the table and the demonstrator’s emotion associated with them were counterbalanced among conditions (Fig. 4).

Control conditions. Before starting with the experimental conditions (see above), control conditions were administered to evaluate object preferences, robot gender recognition, and emotion labeling. More specifically, prior to carrying out the experimental conditions, the participants had to express their liking of the two objects, as well as indicate the gender of the robot. Additionally, to ensure that the demonstrators’ expressed emotions were clearly recognizable, participants were asked to view pictures of the human and robot demonstrators while expressing emotions of joy and disgust. Participants might choose from six different emotions, specifically: joy, anger, surprise, sadness, disgust, and fear.

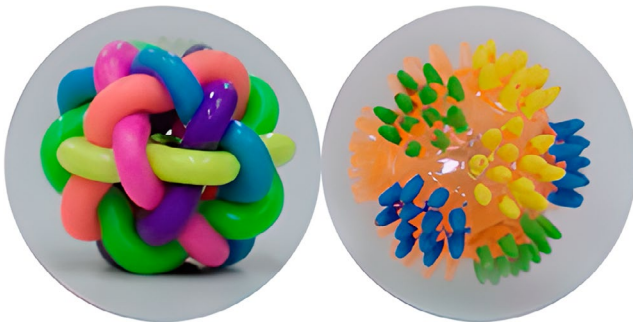


Fig. 4 Unfamiliar objects presented in the Shared Knowledge task: object **A** on the left; object **B** on the right

2.2.3 Correlated Assessments

Besides the Shared Knowledge task, the protocol included the Attribution of Mental States (AMS), the Reflective Functioning Questionnaire (RFQ), and a short version of the Attachment Style Questionnaire.

The Attribution of Mental States questionnaire (AMS-Q) [46]. AMS-Q is a 23-item questionnaire that evaluates the attribution of mental and sensory states to pictures of a human stimulus (female or male). The tool measures the degree of mental anthropomorphization of the non-human agents (e.g., animals, robots, inanimate objects, paranormal entities, and even God) by comparing the scores obtained from the human stimulus with those obtained from the non-human stimuli. The AMS-Q consists of three subscales: AMS-NP, which reflects the attribution of epistemic mental states (e.g., beliefs, thoughts, inferences), well-being states, and positive emotions; AMS-N, which includes the attribution of mental states that belong to the semantic field of deception (e.g., tell a lie) and negative emotions; and AMS-S which refers to sensory states (e.g., hear, smell). Participants were asked to rate each item according to a 5-point Likert scale ranging from 1 (*No, not at all*) to 5 (*Yes, very much*). The scoring was calculated by averaging the items for each factor. The AMS-Q has been used in previous work [31, 32, 38, 49] and has been shown to be a consistent measure in the attribution of mental states to both humans and robots. The questionnaire was administered twice with a human and a robot image as stimuli in random order. The reliability of the scale was excellent, with a Cronbach alpha coefficient reported of 0.96 and 0.91, respectively.

Reflective Functioning Questionnaire (RFQ) [45]. The brief version of the RFQ comprises two subscales, assessing the degrees of uncertainty (RFQ_U) and certainty (RFQ_C) about mental states. The Italian brief version [47] is composed of 8 items that are scored by the participant on a 7-point Likert scale (ranging from “completely disagree” to “completely agree”). As a result, the low agreement reflects hypermentalizing, while some agreement reflects adaptive levels of certainty about mental states. The internal consistency of the sample test was acceptable, with a Cronbach alpha coefficient reported of 0.71.

The Attachment Style Questionnaire (ASQ) [44, 48]. ASQ is a 40-item self-report questionnaire, designed to measure five dimensions of adult attachment: Confidence in Self and Others (8 items), Discomfort with Closeness (10 items), Relationships as Secondary (7 items), Need for Approval (7 items), and Preoccupation with Relationships (8 items). Each item is rated on a 6-point scale, ranging from 1 (totally disagree) to 6 (totally agree). In the current study, we administered three out of the five subscales were administered, i.e., Trust, which reflects a secure attachment

orientation; Need for Approval, which reflects respondents' need for acceptance and confirmation from others; and Concern for Relationships, which involves an anxious and dependent approach to relationships. In the current study, the Cronbach alpha coefficient was 0.80, showing good internal consistency of the sample test.

2.3 Data Analysis

A General Linear Model (GLM) analysis was performed to assess the impact of context (communicative vs. non-communicative) and requester (same requester, different requester, another agent) on participants' choice of target objects under two conditions: human-demonstrator condition and robot-demonstrator condition. The proportion of congruent responses, i.e., when participants chose to share the object with positive valence, was the dependent variable. The participants' object preference was then used as a covariate in the $2 \times 3 \times 2$ repeated measures GLM, with 2 levels of *context* (communicative, non-communicative), 3 levels of *requester* (same identity; different identity, other identity) as the within-subject factors; and 2 levels of *demonstrator* (human, robot) as the between-subject factor. The Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity ($p < .05$). Post-hoc comparisons were Bonferroni corrected.

Table 2 Proportions in emotion labelling in the emotion recognition task

Joy	N (%)	T-test (<i>p</i>)
Human 1	168 (87.0%)	< .001
Human 2	175 (90.7%)	< .001
QT Robot	188 (97.4%)	< .001
Disgust	N (%)	T-test (<i>p</i>)
Human 1	171 (88.6%)	< .001
Human 2	153 (79.3%)	< .001
QT Robot	178 (92.2%)	< .001
Fear	N (%)	T-test (<i>p</i>)
Human 1	2 (1.0%)	< .001
Human 2	0 (0%)	< .001
QT Robot	3 (1.6%)	< .001
Sadness	N (%)	T-test (<i>p</i>)
Human 1	3 (1.6%)	< .001
Human 2	1 (0.5%)	< .001
QT Robot	7 (3.6%)	< .001
Surprise	N (%)	T-test (<i>p</i>)
Human 1	17 (8.8%)	< .001
Human 2	16 (8.3%)	< .001
QT Robot	5 (2.6%)	< .001
Anger	N (%)	T-test (<i>p</i>)
Human 1	15 (7.8%)	< .001
Human 2	39 (20.2%)	< .001
QT Robot	1 (0.5%)	< .001

Furthermore, a GLM analysis was used to assess whether the participants discriminated between the human and robot's mental states, whereas independent binomial logistic regressions were carried out to assess possible predictive effects of the participants' reflective functioning skills and attachment style on responses in the Shared Knowledge task.

3 Results

3.1 Object Preference

Participants were asked to express their liking for the two objects. They showed a significant preference for object A: 48.2% responded "like" for object A vs. 14.5% for object B. For object B response distribution mainly fell between "neutral" (29.5%) or "like a little" (33.7%). Object preference was controlled both by design through randomization of objects' location and associated emotions (see Methods above), and by statistics, i.e., by including the variable "object preference" as a covariate in the GLM analysis carried out to examine participants' responses in the Shared Knowledge task.

3.2 Robot Gender

As the actors involved in the study were men, we assessed the participants' perception of the robot's gender to ensure the gender match. Overall, the robot (QT robot) employed in the study as the demonstrator has been correctly identified as male (78.8%). About 20% answered "don't know", showing that some people do not consider the robot to belong to a specific gender. These data are in line with the results of a preliminary pilot study also evaluating people's identification of the robots' gender.

3.3 Emotion Recognition

The emotions expressed by humans and robots were correctly recognized by most participants. Those who did not correctly name the emotions still correctly indicated the positive or negative valence of the observed emotion (i.e., joy, disgust, fear, sadness, surprise, and anger). The proportions of emotion recognition are given in Table 2.

3.4 Main Analysis: Shared Knowledge Task

The binomial analysis first revealed that the proportion of congruent responses was significantly above the chance level for all conditions ($p < .001$), indicating that the object with positive valence was more likely to be shared with the

requesters independently of context (communicative vs. non-communicative), requester’s identity (same, different, other), and demonstrator’s genus (human, robot). Additionally, by introducing “object preference” as a covariate in the GLM analysis below, the results further showed no substantial correlations between the participants’ responses in the Shared Knowledge task and object preference ($p > .05$), thus indicating that object preference did not influence participants’ choices (see also analysis of covariates below). Values of the GLM with and without covariate are reported in Table 3.

The main results of the GLM analysis related to the Shared Knowledge task showed a significant interaction between demonstrator and context, $F(1, 189)=38,81$, $p < .001$, $partial-\eta^2 = 0.17$, $\delta=1$, and demonstrator and requester, $F(2, 188)=0.85$, $p < .05$, $partial-\eta^2=0.06$, $\delta=90$, suggesting a difference in the effectiveness of the ostensive cues and in the role played by human and robot in the processes of shared knowledge. Also, a significant three-way interaction was found between context, demonstrator, and requester, $F(2, 188)=38,81$, $p < .05$, $partial-\eta^2=0.05$, $\delta=0.80$.

First, under the human-demonstrator condition, pairwise comparisons showed that participants were more likely to share the target object (i.e., the positively valenced object)

with the same person that acted as the demonstrator in the non-communicative context than in the communicative context, $Mdiff=0.17$, $SE=0.05$, $p < .001$. These data partially support the results of the original work ([15]; Fig. 6). Also, a significant difference was found between sharing with the same person acting both as demonstrator and requester rather than with different human acting as a requester in the non-communicative context, $Mdiff=0.13$, $SE=0.05$, $p < .05$.

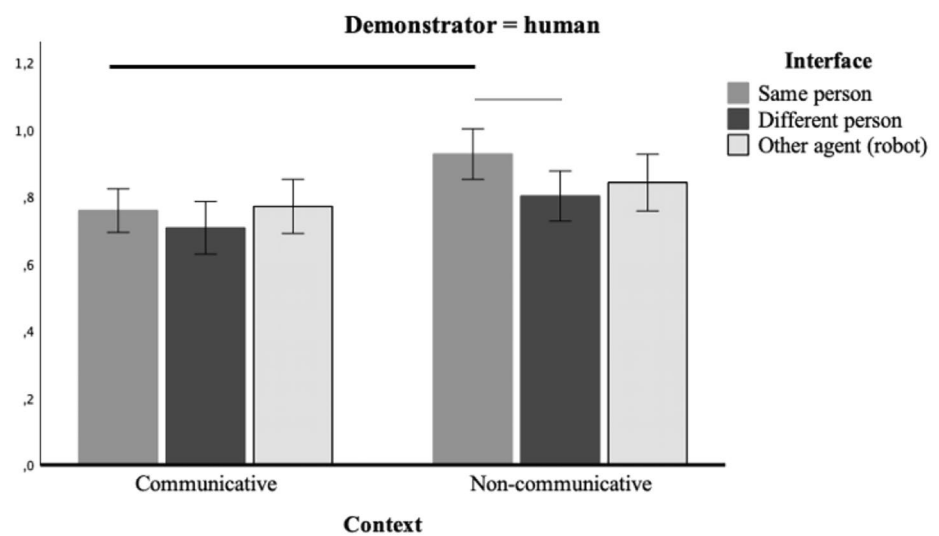
Secondly, under the robot-demonstrator condition, pairwise comparisons showed that participants were more likely to share the target object with the same robot that acted as a demonstrator in a communicative than a non-communicative context, $Mdiff=0.26$, $SE=0.05$, $p < .001$. Furthermore, the target object was more likely to be shared with a human in the communicative than non-communicative context, $Mdiff=0.13$, $SE=0.05$, $p < .05$. Within the communicative context, the target object was more likely to be shared with the same robot that acted as demonstrator than with the human, $Mdiff=0.14$, $SE=0.05$, $p < .05$; whereas, in the non-communicative contexts, it was more likely to be shared with the other robot than with both the same robot, $Mdiff=0.16$, $SE=0.05$, $p < .05$, and the human, $Mdiff=0.17$, $SE=0.05$, $p < .05$ (Fig. 5).

Table 3 Scores of robot demonstrator and human demonstrator conditions with and without object preference as a covariate

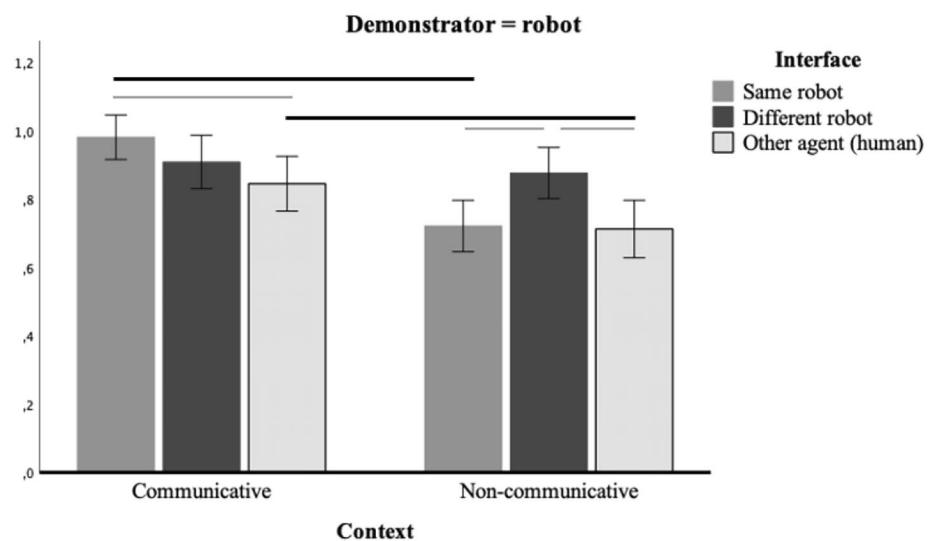
Demonstrator	Context	Requester	Pairwise comparisons without covariate		Pairwise comparisons with covariate		
			<i>Mdiff</i>	<i>SE</i>	<i>Mdiff</i>	<i>SE</i>	
Robot	Communicative	Same robot	Different robot	0.07	0.05	0.07	0.05
			Human	0.13*	0.05	0.14*	0.05
		Different robot	Same robot	-0.07	0.05	-0.07	0.05
			Human	0.06	0.05	0.06	0.05
		Human	Same- Robot	-0.13*	0.05	-0.14*	0.05
			Different robot	-0.06	0.05	-0.06	0.05
	Non-Communicative	Same robot	Different robot	-0.16*	0.05	-0.16*	0.05
			Human	0.01	0.05	0.01	0.05
		Different robot	Same robot	0.16*	0.05	0.16*	0.05
			Human	0.17*	0.05	0.16*	0.05
		Human	Same robot	-0.01	0.05	-0.01	0.05
			Different robot	-0.17*	0.05	-0.17*	0.05
Human	Communicative	Same person	Different person	0.05	0.05	0.05	0.05
			Robot	-0.01	0.05	-0.01	0.05
		Different person	Same person	-0.05	0.05	-0.05	0.05
			Robot	-0.06	0.05	-0.06	0.05
		Robot	Same person	0.01	0.05	0.01	0.05
			Different person	0.06	0.05	0.06	0.05
	Non-communicative	Same person	Different person	0.13*	0.05	0.13*	0.05
			Robot	0.08	0.05	0.09	0.05
		Different person	Same person	-0.13*	0.05	-0.13*	0.05
			Robot	-0.04	0.05	-0.04	0.05
		Robot	Same person	-0.08	0.05	-0.09	0.05
			Different person	0.04	0.05	0.04	0.05

Also, in the communicative context, pairwise comparisons showed that participants were more likely to share the target object with the robot (either the same or different) when the robot played as the demonstrator than with the human (either the same or different), $M_{diff}=0.22$, $SE=0.05$, $p<.001$; $M_{diff}=0.20$, $SE=0.06$, $p<.001$, respectively. In contrast, in the non-communicative context, participants were more likely to share the target object with the same human as the requester than with the same robot acting as both demonstrator and requester, $M_{diff}=0.21$, $SE=0.05$, $p<.001$, and to share with the human when the robot was the demonstrator and with a human when the robot was the demonstrator than with the robot when the demonstrator was human.

Fig. 5 Interaction effect divided by the demonstrator's agency (human, robot), highlighting the differences between requesters within and between communicative contexts



Covariates in the model are estimated at the following values: object_preference_A = 3.39, object_preference_B = 2.38



Covariates in the model are estimated at the following values: object_preference_A = 3.39, object_preference_B = 2.38

3.5 AMS-Q

To assess whether the human and robot were perceived as distinct entities from a mental content and sensory attributes perspective, a 2×2 GLM analysis was carried out, with two levels of AMS (mental states, sensorial states) and two levels of agent (human, robot). The results showed a main effect of AMS, $F(1, 191)=58.70$, $p<.001$, $partial-\eta^2=0.24$, $\delta=1$, indicating a greater attribution of sensory than mental states, and a main effect of agent, $F(1, 191)=2548.12$, $p<.001$, $partial-\eta^2=$

0.93 , $\delta=1$, indicating that the robot scored significantly lower than the human in states attribution. A significant interaction between AMS and agent, $F(1, 191)=63.8$,

$p < .001$, $partial-\eta^2 = 0.25$, $\delta = 1$, also showed that – for the robot – sensory states attribution was substantially greater than mental states attribution, $M_{diff} = 0.43$, $SE = 0.05$, $p < .001$. This difference was not present for human, $p > .05$.

3.6 Logistic Regression

Before running the binomial logistic regressions to assess possible predictive effects on participants' responses, we carried out Pearson's correlation analysis examining the relations between the Shared Knowledge task and participants' reflective functioning skills and attachment style. The analysis yielded a relation between the non-communicative/other-agent condition when the robot was the demonstrator, and Certainty (RFQ_C) about the mental states of self and others, $r(97) = 0.22$, $p < 0.01$. A significant negative relationship was found between the non-communicative context/other-agent condition when the human was the demonstrator and the subscale of ASQ, namely, Need for Approval, $r(96) = -0.27$, $p < .01$; while, in the communicative context, the specular condition was moderately associated to the Concern about relationships of ASQ, $r(96) = 0.21$, $p < .05$.

Based on these results, we carried out binomial logistic regression on three conditions of the Shared Knowledge task with Bonferroni adjustments (p values < 0.016 considered significant) that

correlated with the RFQ and ASQ. The model included five independent variables (RFQ_C, RFQ_U, ASQ-Trust, ASQ-Need for Approval, and ASQ-Concern about relationship). Linearity of the continuous variables with respect to the logit of the dependent variable was assessed via the Box-Tidwell procedure: all continuous independent variables were found to be linearly related to the logit of the dependent variable. We ran three independent logistic regression models for each condition, to outline possible predictive effects of the participants' reflective functioning skills and attachment style on responses in the Shared Knowledge task.

1) *Robot demonstrator – human requester, non-communicative context.* When the robot was the demonstrator, in the non-communicative context/other agent (i.e., when the requester was a human) condition, there was one standardized residual with a value of -3.28 standard deviations and the associated case was deleted from the analysis. Thus, binomial regression was performed again. The full model containing all predictors was statistically significant, $\chi^2(5, N=96)$, 14.43, $p = .013$, indicating that the model is able to distinguish between those who gave a coherent answer versus those who gave an incoherent answer. The model correctly classified

76.0% of cases, indicating the correct identification of the coherent answer (i.e., the positive valence of the object conveyed by the robot). As shown in Table 4, only two of the independent variables made a statistically significant contribution to the model: RFQ_C and ASQ-Need for approval.

2) *Human demonstrator – robot requester, non-communicative context.* Similarly, when the human was the demonstrator, in the non-communicative context-other agent (i.e., when the requester was a robot) condition, there were two standardized residuals with a value of -4.59 and -3.88 standard deviations, which were discarded. Once again, the binomial regression was performed, and the assumption of linearity was not violated. The logistic regression model was statistically significant $\chi^2(5, N=94)$, 18.32, $p = .003$. The model correctly classified 88.3% of cases, indicating the correct identification of the coherent answer (i.e., the positive valence of the object conveyed by the human). In this condition, the predictor variable of sharing the positive valence conveyed by the human was the Attachment Style Questionnaire's need for approval, recording an odds ratio, $\text{Exp}(B)$, of 0.77 (Table 4).

3) *Human demonstrator – robot requester, communicative context.* Lastly, when the demonstrator was acted by a human, in the communicative context-other agent (i.e., when the requester was a robot) condition, three standardized residuals with values above 2.5 were eliminated as these were clear outliers. The new binomial logistic regression showed that the full model containing all predictors was statistically significant, $\chi^2(5, N=93)$, 14.25, $p = .014$. The model correctly classified 82.8% of cases, indicating the correct identification of the coherent answer (i.e., the positive valence of the object conveyed by the human). Also in this condition, two of the independent variables made a statistically significant contribution to the model: ASQ-Need for approval and ASQ-Concern about relationships (Table 4).

4 Discussion

The present study aimed to assess whether the Shared Knowledge assumption, driven by Natural Pedagogy Theory, is a general feature of human communication and therefore persists into adulthood. Additionally, this study aimed at better understanding human-robot interaction, and, for this purpose, we investigated whether the robot, when using ostensive signals, prepares the addressee for the intent to communicate generalizable information. To this end, we developed a

Table 4 Logistic regression predicting the likelihood of sharing object

		B	SE	Wald	df	p	Exp(B)	95% CI for EXP(B)		
									Lower	Upper
Robot_dem, Non-comm, Human_req	ASQ_NeedApp	-0.15	0.06	5.93	1	0.015	0.86	0.77	0.97	
	RFQ_C	-0.18	0.07	6.15	1	0.013	0.83	0.72	0.96	
Human_dem, Non-comm, Robot_req	ASQ_NeedApp	-0.26	0.9	8.76	1	0.003	0.77	0.65	0.92	
Human_dem, comm, Robot_req	ASQ_NeedApp	-0.12	0.06	3.71	1	0.05	0.89	0.79	1	
	ASQ_CAR	0.17	0.06	7.37	1	0.007	1.18	1.05	1.33	

Note: *ASQ_NeedApp* = Need for Approval (*ASQ*); *ASQ_CAR* = Concern about Relationships (*ASQ*); *RFQ_C* = Certainty about the mental states of self and others

paradigm inspired by Egyed et al.'s work [15], in which participants, after having observed the agent's positive or negative emotions toward two objects, had to decide which object to be shared with a requester. Generally, the results of the present study showed (1) that the process of Shared Knowledge previously evaluated in children persists into adulthood, and (2) a fairly different pattern of behavior, when the demonstrator was human or robot, i.e., the positively valenced object (target object), was shared differently depending on whether the demonstrator or requester was a human or a robot, as well as on the contexts in which the demonstrator delivered the information (i.e., communicative vs. non-communicative).

4.1 Human-Demonstrator Condition

Under the human-demonstrator condition, the results generally support the conclusions of the original work [15], thus generalizing the paradigm carried out in presence of young children to adults. Specifically, we found that the target object was more likely to be shared with the same person than with another in the non-communicative contexts. Additionally, we further found that the tendency to share with the same person in the non-communicative context is even greater than sharing with the same person in the communicative context (this condition was absent in the original work). According to Egyed and colleagues [15], our data would support the Shared Knowledge assumption of Natural Pedagogy Theory, which assumes that in a non-communicative context, children do not generalize the agent-specific attributions as applying to other individuals. Moreover, in the absence of ostensive cues, children assign a person-centered interpretation, whereby they interpret the received emotional information as valid only in relation to the referent in the current episodic situation. This idea is further supported by data from our study showing a lack of differences in object-sharing between requesters specifically in the communicative context: under the human-demonstrator condition, the target object was almost equally shared with the *same* person, with a *different* person, or even with the robot. As postulated by Natural Pedagogy Theory, communicative

contexts place the addressee in an attentional state and prepare him or her to receive a subsequent communication containing information specifically relevant to him or her that should be remembered and encoded with other knowledge relevant to social situations [7, 9, 15]. Crucially, the ostensive cues that typically lead the infant to feel recognized as a subject [9], appear to exert their effects even on adults. Our results, consistent with Marno and colleagues' study [16], seem to confirm that ostensive signals have effects beyond simple attentional arousal but prepare the addressee for generalizable knowledge in adult communication as well. Additionally, ostensive cues facilitate the relationship between demonstrator and requester as they trigger the epistemic trust that allows the addressee of object-directed emotion to trust the authenticity of the shared information. Thus, it is plausible to claim that ostensive communication triggers a sense of trust in the person conveying the information as a benevolent, cooperative, and reliable source of cultural information.

4.2 Robot-Demonstrator Condition

The results paint a quite different picture when the robot was the demonstrator. Opposite to what is described above, the target object was more likely to be shared with the same robot in the communicative context than in the non-communicative context. In the communicative context, the target object was shared equally with the *same* robot and with a *different* robot and was less likely to be shared with the human requester. When the robot displayed an object-direct emotion, the human requester appeared as unprivileged as participants were less inclined to consider information received as generalizable to humans, but conversely applicable to any other robots. These results seem to suggest that the information the robot conveys might be considered "robot-specific". A possible explanation for this finding lies in the Theory of Natural Pedagogy that the expectation of learning generalizable knowledge is driven by members of the same social group [6, 7, 50] and, as a matter of fact, robots are not perceived as belonging to the same social group as humans (as evidenced by the data of the AMS-Q).

A similar tendency was observed in the non-communicative context, in which the target object was less shared with the human requester. In contrast to the human-demonstrator condition, in the non-communicative context, participants did not generalize the object positively valenced by the robot to the *same* robot - as postulated by Natural Pedagogy Theory - but rather generalized the target object to the *different* robot. These findings suggest that when dealing with robots, Natural Pedagogy assumptions on the genericity bias are no longer valid. Some other processes would be – as per our data – in place.

Overall, the data collected in the robot-demonstrator conditions bring out the crucial role of ostensive cues in human-robot interactions. We asked whether and how robotic ostensive cues may influence interactions with humans. Although participants were not inclined to generalize the object-directed emotion displayed by the robot to the human requester, in the communicative context participants “listened to the robot” by paying attention to the expressed preference and sharing it with the *same* or *different* robot. In contrast, in the non-communicative contexts, participants, not being ostensibly engaged by the robot demonstrator, tended to share more of the target object with the different requesting robot, which importantly always began the interaction by communicatively addressing the participants. It is, therefore, possible to hypothesize that when a robot does not communicate ostensively, people may not attribute a communicative intention to the robot and, consequently, do not consider it a communicative partner [51]. The human-robot interaction apparently relies on a fundamental feature of human communication, namely the attribution of a communicative intention [8], that is generally afforded by ostensive cues (e.g., direct eye contact, direct speech, calling one’s own name, or contingent response) [20, 52]. When robots communicate ostensively, the addressee attributes a communicative intention to the robot [53]: verbalizations and gaze behavior facilitate the interpretation of the robot’s actions as communicative acts specifically directed at the addressee, leading the addressee to turn the robot into a communicative partner. By attributing communicative intentions, the addressee may consider the information and the communicator’s beliefs, views, and attitudes toward an object, even if conveyed by a robot. Our results are consistent with previous studies with children [23, 51, 53], in which the role of robotic ostensive cues has been found to be important, e.g., a robot that displays ostensive signals can facilitate the acquisition of information and learning [23], and intention attribution (e.g., [53]). Furthermore, according to the Natural Pedagogy theory, communicative signals play a primary role in conveying relevant information because the addressee recognizes the agent’s actions as communicative

acts, understands the intention to communicate, and feels involved as the recipient of the communication [7, 50].

4.3 Attachment and Mentalization Ability in the Shared Knowledge Task

Participants’ attachment styles and specific reflective functioning processes were evaluated for explaining the participants’ behavior and choices in the Shared Knowledge task. Regression analyses showed that participants’ attachment style and reflective functioning predicted their responses in the Shared Knowledge task. Ostensive cues, such as eye contact, accurate turn-taking, and appropriate contingent responsiveness (in time, tone, and content), used by the responsive caregiver to communicate consistent and clear emotional responses, increase the likelihood of a secure child-parent attachment. At least in infancy, ostensive cues can be viewed from a developmental perspective because they trigger a basic epistemic trust in the caregiver as a benevolent, cooperative, and reliable source of cultural information that facilitates the rapid learning of shared knowledge without the need to critically scrutinize its validity or relevance [12, 54]. Conversely, insecure attachment creates epistemic uncertainty [9] and the child constantly tests the trustworthiness of the information delivered by the caregiver. In this sense, attachment bonds serve as a guarantee of the authenticity of knowledge. Our data showed that insecure attachment – resulting in the need for approval and the attitudes of anxiety and dependence on relationships [48] – predicted a less eagerness to share with the robot in the conditions in which the human was the demonstrator. This was independent of context (communicative or non-communicative). Also, those more in need of approval were less likely to share with the human in the conditions in which the robot was the demonstrator, this time only in the non-communicative context. The data first inform us that, generally, when the demonstrator and requester are of different entities, a greater need for approval results in less probability to share. A fine-grained analysis of the data further suggests that this is especially true when the relationship is first established with a human (demonstrator), and one must subsequently share with a robot. This dynamic (i.e., sharing less with the robot when the human was the demonstrator) may tentatively suggest a human-centric approach to relationships. The latter observation is particularly relevant as Csibra and Gergely emphasize that Shared Knowledge should be protected from deliberate distortion by individuals who do not share the same “genetic material”. Indeed, when first engaged by a human, those most in need of approval could be regarded as either skeptical of the relationship with the robot or not want to “betray” the newly constructed relationship with the human demonstrator by tending to share

with the other genus the human's least favorite object. This would also be the reason that those who are most in need of approval and who were initially approached by a robot were less likely to share the target object with the human, especially in conditions where the robot did not engage them via an ostensive cue.

We also found a negative predictive power of the reflective functioning subscale *certainty of one's own and others' mental states* and sharing the target object with a human when the robot demonstrator did not engage participants via an ostensive cue. That is, the greater the participants' certainty with respect to their own and others' mental contents, the lower the likelihood of sharing with the other genus in a non-communicative context. This could possibly mean that participants cannot generalize the information delivered by an informant whose mind is, by its nature, opaque. These would, overall, be consistent with the idea proposed above that in the robot non-communicative context condition, in which participants were less likely to share the target object with a human than with a different robot. The regression data enrich this observation by suggesting that this phenomenon is mediated by confidence in one's own mental abilities, which may be better applied when the mind to be read is human rather than robotic, whose content is unknown. Put another way, good mentalistic skills need to be nurtured by an understanding of the other's mind for "safe" sharing to occur, even more so if the genus acting as the mediator of the relationship (robot) has an opaque mental content and is not conducive to relational engagement with communicative cues.

5 Concluding Remarks and Limitations

In line with the theory of Natural Pedagogy, ostensive cues play a primary role both in human-human interaction and in human-robot interaction and make it possible to efficiently convey information because the addressee assigns to the human or robot demonstrator's actions a communicative intent. Ostensive cues seem to generate the genericity bias in adults as well, namely, the information conveyed in a communicative context is interpreted as generic and extendable to other individuals; this is not the case in non-communicative contexts where information is considered episodic and personal dispositions. In sum, we have demonstrated that, just like in infants, ostensive cues modulate the attention and information encoding as an object- or a person-centered in adults as well; potentially configuring the Shared Knowledge assumption as an inherent part of human communication rather than specific to certain age groups.

Our results further suggest that also non-human ostensive cues elicited a similar attribution of a communicative

intention to the addresser. These findings provide new evidence that robotic ostensive cues play a distinct role in human-robot interaction, allowing the robot to become an effective communicative partner. The crucial point is that the robot must first be considered a social agent with a relational intention. If the communication is not introduced by ostensive cues, the addressee does not consider the robot a communicative partner and does not pay attention to the information the robot wants to convey; hence, the genericity bias is not applied any longer. Moreover, the knowledge is shared by the members of the same social group and, as the AMS-Q data shows, robots and humans belong to two different genera. These differences outline an "inter-agent" discriminative attitude toward the robot. This is also evident from the fact that participants tend to attribute more inter-individual differences between people than between robots, which are perceived to be fundamentally the same precisely because what one robot likes is generalizable to any other robot (but not to humans).

Although our study contributes to human-robot interaction research and also provides input for practical use, some limitations need to be considered. Firstly, we did not investigate participants' familiarity with robots. Secondly, the Shared Knowledge tasks were administered online. We were fully aware that if we had shown a real robot, the physical embodiment might have produced an enhanced effect for participants and eased the influence of the robot's ostensive cues on the affective evaluation of the object and on sharing it. It is important to note that physical and social embodiment are inherently interconnected. From a fundamental perspective, physical embodiment refers to the space occupied by the robot and its ability to move and perceive its surrounding environment. When a second agent is introduced, social interaction also comes into play, even if there is no direct communication between the two parties. It might be worth exploring this in a more ecological context. That said, the fact that our results replicate the findings of the original study allows us to predict with a good degree of certainty the persistence of Shared Knowledge in adulthood and the effectiveness of the robot's ostensive cue if the paradigm of the present study were administered in the presence. Future studies are needed to examine the extent of ostensive cues in human-robot interaction and also to clarify whether the effects found will persist even when people are already familiar with a robot and perceive it within the relationship, such as a household robot. Such research may guide future directions for humanoid robot design in the field of social robotics and lead to new learning strategies.

Acknowledgements We thank Ermanno Bonazzi and Nicolò Di Dio for helping in the preparation and production of the experimental videos.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by CDD, LM, and FM. The first draft of the manuscript was written by CDD and LM, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Further considerations.

•Manuscript has been “spellchecked” and “grammar checked”.

•All references mentioned in the Reference List are cited in the text, and vice versa.

•Permission has been obtained for use of copyrighted material from other sources (including the Internet).

Funding Open access funding provided by Università Cattolica del Sacro Cuore within the CRUI-CARE Agreement.

Data Availability The dataset generated during and/or analyzed during the current study is available from the corresponding author on reasonable request.

Statements and Declarations

Conflict of Interest The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval All procedures performed in this study involving human participants were in accordance with the 1964 Declaration of Helsinki and its subsequent revisions or comparable ethical standards. The study was approved by the Ethics Committee of the Università Cattolica del Sacro Cuore of Milan, Department of Psychology (date: May 27, 2021; No. 49–21).

Informed Consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Moses LJ, Baldwin DA, Rosicky JG, Tidball G (2001) Evidence for Referential understanding in the Emotions Domain at twelve and eighteen months. *Child Dev* 72:718–735. <https://doi.org/10.1111/1467-8624.00311>
- Mumme DL, Fernald A (2003) The infant as onlooker: learning from emotional reactions observed in a television scenario. *Child Dev* 74:221–237. <https://doi.org/10.1111/1467-8624.00532>
- Baldwin DA, Moses LJ (1996) The Ontogeny of Social Information gathering. <https://doi.org/10.2307/1131601>. *Child Development* 67:1915
- Tomasello M (1999) The human adaptation for culture. *Annu Rev Anthropol* 28:509–529. <https://doi.org/10.1146/annurev.anthro.28.1.509>
- Isernia S, Baglio F, d'Arma A et al (2019) Social mind and long-lasting disease: focus on affective and cognitive theory of mind in multiple sclerosis. *Front Psychol* 10:218. <https://doi.org/10.3389/fpsyg.2019.00218>
- Csibra G, György G (2006) Social learning and social cognition: the case for pedagogy. *Atten Perform* 21:249–274
- Csibra G, Gergely G (2009) Natural pedagogy. *Trends Cogn Sci* 13:148–153. <https://doi.org/10.1016/j.tics.2009.01.005>
- Grice P (1991) *Studies in the way of words*. Harvard University Press, Cambridge, MA
- Fonagy P, Allison E (2014) The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy* 51:372–380. <https://doi.org/10.1037/a0036505>
- Luyten P, Campbell C, Allison E, Fonagy P (2020) The Mentalizing Approach to Psychopathology: state of the art and future directions. *Ann Rev Clin Psychol* 16:297–325. <https://doi.org/10.1146/annurev-clinpsy-071919-015355>
- Fisher S, Guralnik T, Fonagy P, Zilcha-Mano S (2021) Let's face it: video conferencing psychotherapy requires the extensive use of ostensive cues. *Counselling Psychol Q* 34:508–524. <https://doi.org/10.1080/09515070.2020.1777535>
- Fonagy P, Gergely G, Target M (2007) The parent?infant dyad and the construction of the subjective self. *J Child Psychol & Psychiat* 48:288–328. <https://doi.org/10.1111/j.1469-7610.2007.01727.x>
- Baldwin DA (1993) Early referential understanding: Infants' ability to recognize referential acts for what they are. *Dev Psychol* 29:832–843. <https://doi.org/10.1037/0012-1649.29.5.832>
- Flom R, Lee K, Muir D (2007) Gaze following. *Its Development and Significance*
- Egyed K, Király I, Gergely G (2013) Communicating Shared Knowledge in Infancy. *Psychol Sci* 24:1348–1353. <https://doi.org/10.1177/0956797612471952>
- Marno H, Davelaar EJ, Csibra G (2014) Nonverbal communicative signals modulate attention to object properties. *J Exp Psychol Hum Percept Perform* 40:752–762. <https://doi.org/10.1037/a0035113>
- Yoon JMD, Johnson MH, Csibra G (2008) Communication-induced memory biases in preverbal infants. *Proc Natl Acad Sci USA* 105:13690–13695. <https://doi.org/10.1073/pnas.0804388105>
- Csibra G (2010) Recognizing communicative intentions in infancy. *Mind & Language* 25:141–168. <https://doi.org/10.1111/j.1468-0017.2009.01384.x>
- Gergely G (2007) Learning 'about' versus learning 'from' other minds: human pedagogy and its implications. In: Carruthers P (ed) *The innate mind: foundations and the future*. Oxford University Press, Usa
- Vorms M (2012) A-not-B errors: testing the limits of natural pedagogy theory. *RevPhilPsych* 3:525–545. <https://doi.org/10.1007/s13164-012-0113-4>
- Okumura Y, Kanakogi Y, Kobayashi T, Itakura S (2020) Ostension affects infant learning more than attention. *Cognition* 195:104082. <https://doi.org/10.1016/j.cognition.2019.104082>
- Senju A, Csibra G (2008) Gaze following in Human Infants depends on communicative signals. *Curr Biol* 18:668–671. <https://doi.org/10.1016/j.cub.2008.03.059>
- Okumura Y, Kanakogi Y, Kanda T et al (2013) Can infants use robot gaze for object learning?: the effect of verbalization. *IS* 14:351–365. <https://doi.org/10.1075/is.14.3.03oku>
- Parise E, Cleveland A, Costabile A, Striano T (2007) Influence of vocal cues on learning about objects in joint attention contexts. *Infant Behav Dev* 30:380–384. <https://doi.org/10.1016/j.infbeh.2006.10.006>

25. Schröder-Pfeifer P, Talia A, Volkert J, Taubner S (2018) Developing an assessment of epistemic trust: a research protocol. <https://doi.org/10.4081/ripppo.2018.330>. ResPsy 21:
26. Träuble B, Bätz J (2014) Shared function knowledge: Infants' attention to function information in communicative contexts. *J Exp Child Psychol* 124:67–77. <https://doi.org/10.1016/j.jecp.2014.01.019>
27. Breazeal C, Dautenhahn K, Kanda T (2016) Social Robotics. In: Siciliano B, Khatib O (eds) Springer Handbook of Robotics. Springer International Publishing, Cham, pp 1935–1972
28. Baxter P, Ashurst E, Read R et al (2017) Robot education peers in a situated primary school study: personalisation promotes child learning. *PLoS ONE* 12:e0178126. <https://doi.org/10.1371/journal.pone.0178126>
29. Belpaeme T, Kennedy J, Ramachandran A et al (2018) Social robots for education: a review. *Sci Robot* 3:eaat5954. <https://doi.org/10.1126/scirobotics.aat5954>
30. Cangelosi A, Schlesinger M (2018) From babies to Robots: the contribution of Developmental Robotics to Developmental psychology. *Child Dev Perspect* 12:183–188. <https://doi.org/10.1111/cdep.12282>
31. Di Dio C, Manzi F, Itakura S et al (2020) It does not Matter who you are: Fairness in Pre-schoolers interacting with human and robotic partners. *Int J of Soc Robotics* 12:1045–1059. <https://doi.org/10.1007/s12369-019-00528-9>
32. Di Dio C, Manzi F, Peretti G et al (2020) Come I bambini pensano alla mente di un robot. Il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali a un agente robotico [How children think about the robot's mind. The role of attachment and theory of mind in the attribution of mental states to a robotic agent]. *Sistemi Intelligenti* 1:41–56
33. Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive Robots as Social Partners and peer tutors for children: a Field Trial. *Hum Comput Interact (Special issues human-robot interaction)* 19:61–84. https://doi.org/10.1207/s15327051hci1901&2_4
34. Kont M, Alimardani M (2020) Engagement and mind perception within Human-Robot Interaction. A Comparison between Elderly and Young Adults
35. Marchetti A, Manzi F, Itakura S, Massaro D (2018) Theory of mind and Humanoid Robots from a Lifespan Perspective. *Z für Psychologie* 226:98–109. <https://doi.org/10.1027/2151-2604/a000326>
36. Di Dio C, Manzi F, Peretti G et al (2020) Shall I trust you? From Child–Robot Interaction to Trusting Relationships. *Front Psychol* 11:469. <https://doi.org/10.3389/fpsyg.2020.00469>
37. Vinanzi S, Patacchiola M, Chella A, Cangelosi A (2019) Would a robot trust you? Developmental robotics model of trust and theory of mind. *Phil Trans R Soc B* 374:20180032. <https://doi.org/10.1098/rstb.2018.0032>
38. Manzi F, Peretti G, Di Dio C et al (2020) A Robot is not worth another: exploring children's Mental State Attribution to different Humanoid Robots. *Front Psychol* 11:2011. <https://doi.org/10.3389/fpsyg.2020.02011>
39. Złotowski J, Proudfoot D, Yogeewaran K, Bartneck C (2015) Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *Int J of Soc Robotics* 7:347–360. <https://doi.org/10.1007/s12369-014-0267-6>
40. Manzi F, Di Dio C, Di Lernia D et al (2021) Can you activate me? From Robots to Human Brain. *Front Robot AI* 8:633514. <https://doi.org/10.3389/frobt.2021.633514>
41. Manzi F, Sorgente A, Massaro D, Cyberpsychology et al (2021) *Behav Social Netw* 24:315–323. <https://doi.org/10.1089/cyber.2020.0161>
42. Manzi F, Ishikawa M, Di Dio C et al (2020) The understanding of congruent and incongruent referential gaze in 17-month-old infants: an eye-tracking study comparing human and robot. *Sci Rep* 10:11918. <https://doi.org/10.1038/s41598-020-69140-6>
43. Chevalier P, Kompatsiari K, Ciardo F, Wykowska A (2020) Examining joint attention with the use of humanoid robots-A new approach to study fundamental mechanisms of social cognition. *Psychon Bull Rev* 27:217–236. <https://doi.org/10.3758/s13423-019-01689-4>
44. Feeney JA, Noller P, Hanrahan M (1994) Assessing adult attachment. Attachment in adults: clinical and developmental perspectives. Guilford Press, New York, NY, US, pp 128–152
45. Fonagy P, Luyten P, Moulton-Perkins A et al (2016) Development and validation of a self-report measure of Mentalizing: the reflective functioning questionnaire. *PLoS ONE* 11:e0158678. <https://doi.org/10.1371/journal.pone.0158678>
46. Miraglia L, Peretti G, Manzi F et al (2023) Development and validation of the attribution of Mental States Questionnaire (AMS-Q): a reference tool for assessing anthropomorphism. *Front Psychol* 14:999921. <https://doi.org/10.3389/fpsyg.2023.999921>
47. Morandotti N, Brondino N, Merelli A et al (2018) The italian version of the reflective functioning questionnaire: validity data for adults and its association with severity of borderline personality disorder. *PLoS ONE* 13:e0206433. <https://doi.org/10.1371/journal.pone.0206433>
48. Fossati A, Feeney JA, Donati D et al (2003) On the dimensionality of the attachment style questionnaire in italian clinical and nonclinical participants. *J Social Personal Relationships* 20:55–79. <https://doi.org/10.1177/02654075030201003>
49. Di Dio C, Isernia S, Ceolaro C et al (2018) Growing up thinking of God's beliefs: theory of mind and ontological knowledge. *SAGE Open* 8:215824401880987. <https://doi.org/10.1177/2158244018809874>
50. Csibra G, Gergely G (2011) Natural pedagogy as evolutionary adaptation. *Phil Trans R Soc B* 366:1149–1157. <https://doi.org/10.1098/rstb.2010.0319>
51. Arita A, Hiraki K, Kanda T, Ishiguro H (2005) Can we talk to robots? Ten-month-old infants expected interactive humanoid robots to be talked to by persons. *Cognition* 95:B49–B57. <https://doi.org/10.1016/j.cognition.2004.08.001>
52. Russell B (1940) *An Inquiry Into meaning and Truth*. Routledge
53. Itakura S, Ishida H, Kanda T et al (2008) How to build an intentional android: Infants' imitation of a Robot's Goal-Directed actions. *Infancy* 13:519–532. <https://doi.org/10.1080/15250000802329503>
54. Corriveau KH, Harris PL, Meins E et al (2009) Young children's trust in their mother's claims: longitudinal links with attachment security in infancy. *Child Dev* 80:750–761. <https://doi.org/10.1111/j.1467-8624.2009.01295.x>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Laura Miraglia is a Ph.D. candidate in the Personal and Educational Sciences, specializing in ‘Person, Development, and Learning’ at the Università Cattolica del Sacro Cuore, Department of Psychology, in Milan. She is a member of the Research Unit on Theory of Mind (UniToM). Her research is dedicated to social cognition, the forms of understanding intentions, and the cognitive processes in order to investigate the psychological mechanisms necessary and sufficient for human-robot interaction, taking into account developmental age and typical and atypical development.

Cinzia Di Dio, Ph.D. is Associate Professor of Developmental Psychology and Educational Psychology, Faculty of Education, at the Università Cattolica del Sacro Cuore, Milan. She is co-director of the Research Unit in Psychology and Robotics in the Lifespan (PsyRoLife) and a member of the Research Unit on Theory of Mind (UniToM), Department of Psychology. She is Early Career Scholars Representative of the International Society for the Study of Behavioral Development (ISSBD). She is on the faculty Board of the Doctor of National Interest in “Artificial Intelligence.” She received her Ph.D. in Neuroscience from the University of Parma under the supervision of Professor Giacomo Rizzolatti.

Federico Manzi, Ph.D. is a PON-Innovation researcher in Developmental Psychology and Educational Psychology at the Faculty of Educational at the Università Cattolica del Sacro Cuore, Milan. He is the co-director of the Research Unit in Psychology and Robotics in the Lifespan (PsyRoLife) and a member of the Research Unit on Theory of Mind (UniToM), Department of Psychology. He earned a double doctoral degree in “Person and Formation Sciences” at the UCSC and in “Human Sciences - Developmental Psychology” at the University of Neuchâtel (Switzerland). He has been a visiting researcher at research centers in Japan, the United Kingdom, and Switzerland. He is a robotics expert member of European projects and on the Committee at various international conferences on robotics. He is the Coordinator of Young Researchers Networks at the Italian Association of Psychology (AIP) and a member of the Executive Committee of Early Career Scholars at the International Society for the Study of Behavioral Development (ISSBD).

Takayuki Kanda is a professor of Informatics at Kyoto University, Japan. He is also a Visiting Group Leader at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. He received his B. Eng, M. Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. He is one of the starting members of the Communication Robots project at ATR. He has developed a communication robot, Robovie, and applied it in daily situations, such as a peer tutor at elementary school and a museum exhibit guide. Main research: human-robot interaction, interactive humanoid robots, and field trials.

Angelo Cangelosi, Ph.D. is a Professor of Machine Learning and Robotics at the University of Manchester (UK) and co-director and founder of the Manchester Centre for Robotics and AI. He was awarded a European Research Council (ERC Advanced) project in 2023. Cangelosi also is Turing Fellow at the Alan Turing Institute London and Visiting Professor at Università Cattolica del Sacro Cuore in Milan and Hohai University. His research interests are in cognitive and developmental robotics, neural networks, language grounding, human-robot interaction and trust, and robot companions for health and social care.

Shoji Itakura, Ph.D. Director of the Center for Baby Science, Doshisha University, Kyoto, Japan. He is Emeritus Professor at the Kyoto University. Main research: Developmental Cybernetics, development of social cognition, theory of mind, prosocial behavior, moral development, synchrony.

Hiroshi Ishiguro received a D. Eng. in systems engineering from Osaka University, Japan in 1991. He is currently a Professor of the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009–) and a Distinguished Professor at Osaka University (2017–). He is also visiting Director (2014–) (group leader: 2002–2013) of Hiroshi Ishiguro Laboratories at the Advanced Telecommunications Research Institute and an ATR fellow. Main research: sensor networks, interactive robotics, and android science.

Davide Massaro, Ph.D. is Full Professor of Developmental Psychology and Educational Psychology at the Università Cattolica del Sacro Cuore, Milan. He is a senior member of the Research Unit on Theory of Mind and serves as Academic Coordinator of the Doctoral Program in Personal and Educational Sciences, as well as co-coordinator of the master’s degree program in Media Education. He is vice president of the Italian Psychology Association (AIP). He is part of the faculty Board of the Doctor of National Interest in “Teaching & Learning Sciences”. After graduating in Psychology, he received his Ph.D. in Developmental and Social Psychology from the Università Cattolica del Sacro Cuore in Milan.

Peter Fonagy, Ph.D. is Head of the Division of Psychology and Language Sciences at UCL; Chief Executive of the Anna Freud National Centre for Children and Families, London; Executive Clinical Director UCL Partners Mental Health Programme and National Clinical Advisor on Children’s Mental Health, NHS England. He is a Visiting Professor at Yale University and held visiting professorships at Baylor College of Medicine, Cambridge University, and Harvard Medical Schools. His clinical interests center on issues of early attachment relationships, social cognition, borderline personality disorder, and violence. A major focus of Professor Fonagy’s contribution has been an innovative research-based dynamic therapeutic approach, called Mentalization-Based Treatment, which was developed in collaboration with a number of clinical sites both in this country and in the US.

Antonella Marchetti, Ph.D. is Full Professor of Developmental Psychology and Educational Psychology, Faculty of Education, and Head of the Department of Psychology at Università Cattolica del Sacro Cuore, Milan. She is Head of the Research Unit on Theory of Mind and Coordinator of the Doctoral School in “Person and Formation Sciences.” Member of the Scientific Committee of the Natural and Artificial Cognition Lab (NAC, University of Naples Federico II) and of the Foundation for Financial and Savings Education (FEduF). Regional Coordinator for Italy and member of the Executive Committee of the International Society for the Study of Behavioural Development (ISSBD). She is a member of the Ethics Advisory Board of the Horizon2020 SOLARIS. She is on the faculty board of the Doctor of National Interest in “Robotics and Intelligent Machines.”