International Journal of Social Robotics
https://doi.org/10.1007/s12369-022-00941-7

Infants' Prediction of Humanoid Robot's Goal-Directed Action

F. Manzi[1,2] · M. Ishikawa[3] · C. Di Dio[1,2] · S. Itakura[3] · T. Kanda[4,5] · H. Ishiguro[5,6] · D. Massaro[1,2] · A. Marchetti[1,2]

Accepted: 20 October 2022

Abstract

Several studies have shown that infants anticipate human goal-directed actions, but not robot's ones. However, the studies focusing on the robot goal-directed actions have mainly analyzed the effect of mechanical arms on infant's attention. To date, the prediction of goal-directed actions in infants has not yet been studied when the agent is a humanoid robot. Given this lack of evidence in infancy research, the present study aims at analyzing infants' action anticipation of both a human's and a humanoid robot's goal-directed action. Data were acquired on thirty 17-month-old infants, watching four video clips, where either a human or a humanoid robot performed a goal-directed action, i.e. reaching a target. Infants looking behavior was measured through the eye-tracking technique. The results showed that infants anticipated the goal-directed action of both the human and the robot and there were no differences in the anticipatory gaze behavior between the two agents. Furthermore, the findings indicated different attentional patterns for the human and the robot, showing a greater attention paid to the robot's face than the human's face. Overall, the results suggest that 17-month-old infants may infer also humanoid robot' underlying action goals.

Keywords Infants · Goal-directed action · Action prediction · Humanoid robot · Robot's face

1 Introduction

The current social scenario is gradually shaping a future in which social robots will become partners of our children from an early age [1, 2]. Among the different types of robots, humanoids are particularly useful in the interaction with children already in the first months of life due to their physical and behavioral human resemblance. It is in fact widely recognized that the robots' human-likeness can elicit the activation of similar psychological mechanisms and processes that underpin interactions between humans [3–7]. From early infancy, the understanding of actions made by others plays a key role in making sense of the social world (for a review see [8]). For example, in the relational dynamic between a child and a caregiver it is important to understand that when the caregiver is moving her/his arm towards a glass of water, it means s/he is about to grasp it, and this represents sense-making of the relational world around the child. Action understanding then gradually develops in the child's ability to also predict the partner's behaviors [9]. Although much debated, many studies have interpreted the prediction of an action with the ascription of intentionality to the partner [10, 11]. Let us now imagine a future in which a humanoid robot takes care of our child: s/he is crying because s/he is thirsty and her/his cup of water is not within her/his reach. The robot could move its arm to grasp the cup and give it to the child. Would our child be able anticipate the robot's action, i.e., take the cup, thus understanding the action made by the robot? To date, studies on action prediction have not yet demonstrated the ability of infants to predict the action

F. Manzi and M. Ishikawa authors are contributed equally to this work.

✉ F. Manzi
federico.manzi@unicatt.it

1 Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

2 Research Unit on Robopsychology in the Lifespan, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

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

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

5 Advanced Telecommunications Research Institute International, IRC/HIL, Keihana Science City, Kyoto, Japan

6 Department of System Innovation, Osaka University, Toyonaka, Japan

Published online: 09 November 2022

of a humanoid robot. Starting from this lack of knowledge, in our work we address two critical questions for the infant-robot interaction debate. First, are infants able to predict the action of a humanoid robot? Second, do infants predict the action of a humanoid robot as accurately as they predict the action of a human?

To answer these questions, we start by drawing our attention to the great body of evidence showing the ability to predict the action of a human within the first year of life [12–21]. The results of these studies have consistently proved the child's ability to correctly predict another human's behavior through various methodologies: looking-time (i.e., Woodward-paradigm) [10, 22, 23], anticipatory looking [16, 24–27], behavioral observation [11, 28, 29], and neuroscientific techniques [30, 31]. In the present study, we adopted the anticipatory looking methodology to analyze infants' action prediction. More specifically, the child's gaze-shift to an object (i.e., the target) is analyzed when an agent moves the arm towards the target and grasps it. Anticipatory looking is identified when the child shifts the gaze to the target before the agent grasps it. Predictive gaze behavior reflects the child's ability to recognize the action as goal-directed [32–34].

Turning now to investigations addressing child human–robot relationship, the anticipatory looking methodology to study infants' action prediction has been also recently used to evaluate infants' ability to predict goal-directed actions performed by non-human agents. The majority of these studies evaluated infants' anticipatory looking by comparing a human arm with a mechanical arm [35, 36]. The findings show that at 10 months infants predict the goals of simple and familiar actions (e.g., grasping a toy) performed by familiar agents (i.e., a human hand) whereas they do not predict simple actions when performed by unfamiliar agents, such as mechanical arms [35, 36]. At 12 months, infants predict the action of an unfamiliar agent, such as a mechanical arm, when it exhibits behavioral agency (e.g., self-propelled movement, equifinality of goal achievement, the ability to produce salient action effects; [9, 37]). With respect to the development of action-anticipation of mechanical agents, a recent study by Adam et al. [38] showed that only at 18 months infants are able to anticipate the action of a mechanical arm. From a developmental perspective, taken together the findings on mechanical arms suggest that infants as young as 12 months of age do not present anticipatory looking (i.e., predict an action) of a mechanical arm, which is an ability that appears to emerge at 18 months. Furthermore, the data suggest that simple actions and familiarity with the agent are important components of action-anticipation.

As far as humanoid robots are concerned, in recent years many studies have highlighted that people tend to interact socially with them and this is particularly pronounced in the first years of life [4, 5, 7, 39, 40]. The few studies on infants' social responses to humanoid robots have mainly focused on the child's disposition to follow the robot's gaze by adopting classical gaze-following paradigms, which largely make use of the eye-tracking technique. These studies have demonstrated that—up to 12 months—infants are less responsive to the gaze direction of a humanoid robot, preferring the human gaze [41–43]. Also, they suggest that the referential understanding of humanoid robot gaze begins to emerge at about 17 months of age [40]. Only at 18 months, infants follow the robot gaze, but if they first witness an interaction between the robot and an adult [44], the adult playing as the mediator of the child-robot interaction. With respect to children's response to the robot action, in toddlers between 24 and 35 months old, Itakura et al. [45] examined the imitation processes of a humanoid robot's targeted actions. The finding showed that children at this age can understand intention underpinning the goal-directed actions of a humanoid robot by imitating the targeted action in particular when the robot engage the child with eye-contact [31]. As a matter of fact, the ability to imitate the actions of a humanoid robot was found from the age of 11 months [46]. In this case, the children had to imitate some simple actions from a human and a robotic model. The results show that, although the action of the human remains more evocative than the action of the robot, both agents succeed in evoking motor imitative processes in two-year-old children.

Taken together, the findings on humanoid robots, although not specifically focused on action prediction, support the idea that, from 12 months onwards, infants recognize some human-like characteristics in humanoid robots even if they prefer the human (i.e., they follow the robot's gaze, but prefer the human gaze; they imitate a robot's actions, but imitate the human's actions more). It is only from the middle of the second year of life that children activate complex responses to the behavior of humanoid robots. Very little is known on the other hand about the child's anticipatory behavior in response to the robot's action. In fact, to our knowledge there are neither data on the ability of infants to anticipate the action of a humanoid robot nor data comparing the anticipation of an action performed either by a humanoid robot or a human.

Therefore, with the aim of tackling this unexplored area, and also drawing on data from the studies described above (in particular, related to the mechanical arm), in the present study we analyzed the anticipatory looking of a simple action (i.e., reaching a toy) in 17-month-old infants as they can anticipate the action of a human as well as a non-human agent. To increase the degree of familiarity between the child and the non-human agent, we used a humanoid robot instead of a mechanical arm.

1.1 Aims and Hypotheses

The aim of the present study is to understand the ability of 17-month-olds to predict a simple action (i.e., reaching a toy) performed by a humanoid robot and if this prediction differs when the same action is performed by a human. To achieve this goal, the anticipatory looking methodology was adopted as informative of the child's action prediction [16, 24, 35, 36, 38]. The sample was identified from studies on (i) the anticipatory looking of simple human actions, showing that by 12 months of age infants anticipate the action, (ii) the anticipatory looking of non-human agents, revealing that before 18 months of age infants fail to anticipate simple actions of a mechanical arm, and (iii) infants' early responses to humanoid robots showing that only from the middle of the second year of life children follow the gaze of the humanoid robot and imitate simple actions.

In the present study, we analyzed infants' anticipatory gaze behavior—i.e., the ability to predict an agent's behavior (DV)—with respect to a simple goal-directed action (i.e., reaching a toy) performed by a human and a humanoid robot (IVs). For the present study, we hypothesize that infants at 17 months would (1) anticipate the human's action, (2) anticipate the humanoid robot's action, and (3) anticipate the human's action faster than the humanoid robot's action.

2 Methods

2.1 Participants

Data were obtained from an initial sample of 32 17-month-old infants. We planned to conduct exploratory analysis for the looking data during watching the goal-directed action, thus we did not conduct power analysis based on a specific effect. We aimed to collect clean data from a larger sample size than the previous eye-tracking study (n = 18) examining infants' attention during observing goal-directed manual grasping actions [33]. Two infants were excluded from the analyses because of technical acquisition errors. Therefore, analysis was carried on data from 30 Japanese infants. The infants were randomly assigned to the two groups as follows: (1) 16 infants for the Human condition ($M_{\text{age}} = 17.25$, $SD = 0.83$); (2) 14 infants for the Robot condition ($M_{\text{age}} = 17.64$, $SD = 1.08$). The infants' parents received a written explanation of the procedure of the study, the measurement items, and provided written informed consent before their infants took part in the study. The Research Ethics Review Board of Department of Psychology, Kyoto University, Japan, approved the experimental protocol (ethical proof number: 28-P12).

2.2 Design and Stimuli

The design of the study was a multifactorial 3 × 3 × 2 mixed-model, with 3 levels of Areas of Interest (*AOIs*: Face, Hand, Object), 3 levels of video *Sequence* (Sequence-1, Sequence-2, Sequence-3) as the within-subject factors, and two levels of *Agent* (Human, Robot) as the between-subject factor.

The stimuli developed for the experiment were 8 video clips, in which a human (a woman; see Fig. 1A–D) or the robot (Robovie robot; see Fig. 2A–D) moved her/its right arm reaching a toy (submarine and rocket; see Figs. 1A–D and 2A–D). Depending on the condition (Human or Robot) to which the infants were randomly assigned, each participant watched 4 video clips (target-submarine on the right, target-submarine on the left, target-rocket on the right, target-rocket on the left; see Figs. 1A–D and 2A–D).

The experimental session consisted of a familiarization phase and a test phase. Depending on the condition (Human or Robot), a familiarization video was administered showing the upper body of the human or robot followed by the presentation of a fixation point coupled with an acoustic signal. The purpose of this initial phase was to familiarize infants with the setting and the agents (human or robot).

Each experimental video lasting 6 s began with a scene (Sequence-1) in which the agent looked straight at the camera (2 s), thus establishing an engagement with the infants. Next (Sequence-2), the agent lifted the hand fixating it (2 s), thus shifting the focus to the hand. The agent then (Sequence-3) moved her/its right arm (2 s) reaching the object (submarine or rocket). The three sequences of all the video are of equal length (2 s each). The human model maintained a neutral facial expression and remained silent throughout the entire sequence. Before each experimental video an object (toy animation) appeared at the centre of the screen accompanied by a tinkling sound to attract the infant's attention. It has to be noted that, in normal reaching behaviour, people tend to look at the target they intend to grasp, as eyes lead the hand in actions [47]. Therefore, to control the effect of action anticipation due to the agent's gaze, in Sequence-2 the actor watched the hand lift from the table and in Sequence-3 she/it simultaneously moved the gaze and the hand towards the target. Through this expedient, we managed to focus on the prediction effect of hand movements [48].

2.3 Procedure and Apparatus

The infants were assessed individually in the presence of their mother in the Developmental Science laboratory at Kyoto University. The infants were randomly assigned to the Human or Robot condition and the presentation order of each condition (human and robot; left or right position of the object) was randomized across infants.

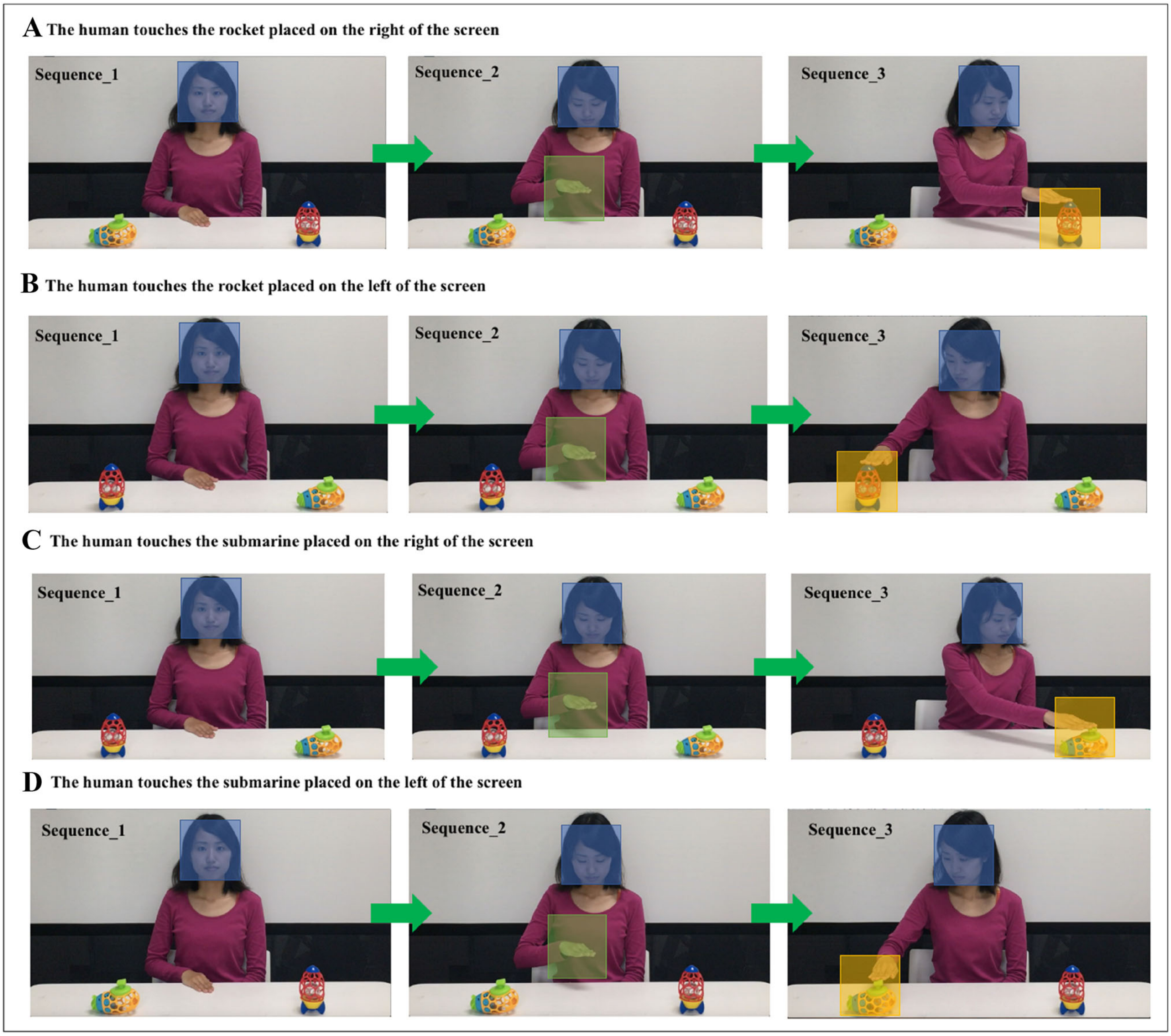

Fig. 1 **A–D** The figure represents the four video stimuli for the Human condition. For all conditions, the three sequences have the same duration. Specifically, in Sequence-1 the actress looked straight into the camera (2 s) and then, Sequence-2, watched the hand lift slightly from the table (2 s). Sequence-3 (2 s) changed according to the target and the location of the two objects: in **A** and **B** the target was the rocket, and this was placed to the left and right of the actress respectively; in **C** and **D** the target was the submarine, and this was placed to the left and right of the actress respectively. AOIs were included for all videos and sequences (in blue the Face, in green the Hand and in yellow the Target)

We used the Tobi T60 (Tobii pro studio, Tobii Technology, Stockholm) to record the infant's eye-gaze. The robot used for the tasks was the humanoid robot Robovie2 (Hiroshi Ishiguro Laboratories). The sampling rate of eye tracking was 60 Hz. The participants were seated on the caregiver's lap approximately 60 cm from the monitor. Prior to recording, a five-point calibration was conducted through the Tobii Studio software 2.2.8 version. More specifically, the calibration stimulus was an animated calibration targets (e.g., moving puppets).

2.4 Data Analysis

For each video, 3 areas of interest (AOIs) were defined as follows: (1) the agent's face, (2) the agent's hand and (3) the target object (the toys that the agent was going to grasp). The AOIs are all equal size. Face and hand have been defined as areas of interest in line with recent studies that demonstrate their importance as early social cues [49]. In addition, each video was divided into three sequences: (1) an initial sequence in which there was no movement or action by the

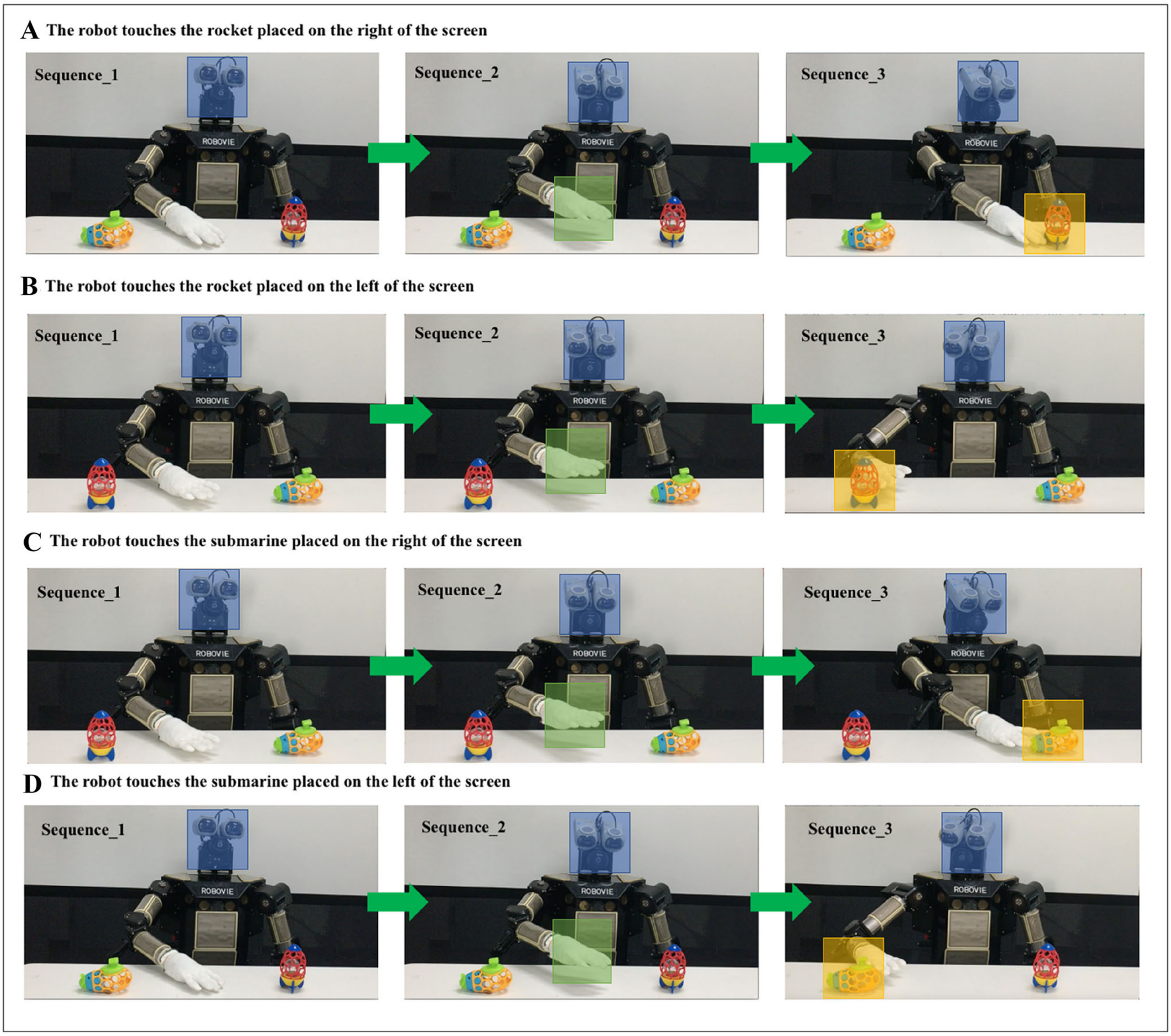

Fig. 2 **A–D** The figure represents the four video stimuli for the Robot condition. For all conditions, the three sequences have the same duration. Specifically, in Sequence-1 the robot looked straight into the camera (2 s) and then, Sequence-2, watched the hand lift slightly from the table (2 s). Sequence-3 (2 s) changed according to the target and the location of the two objects: in **A** and **B** the target was the rocket, and this was placed to the left and right of the robot respectively; in **C** and **D** the target was the submarine, and this was placed to the left and right of the robot respectively. AOIs were included for all videos and sequences (in blue the Face, in green the Hand and in yellow the Target)

protagonist; (2) a sequence that included the movement of the hand and head; (3) a final sequence representing the completion of the action, i.e., touch one of the two toys (submarine or rocket; see Figs. 1A–D and 2A–D).

The dependent variables were the following: (1) time to first fixation on the target in the sequence 3 (Touching Movement) to assess infants' prediction of agent's goal-directed action; (2) the total fixation duration to evaluate infants' general attentional pattern on the different AOIs in the three sequences as a function of agent (human or robot).

To evaluate infants' prediction of the goal-directed action to the targets we adopted the anticipatory gaze methodology [36, 38]. The anticipatory gaze was calculated by subtracting the participants' first time to fixation on the target's AOI from the time when the human's and robot's hand entered the target's AOI (Sequence 3) [36, 38]. The child's gaze was considered predictive (positive score) if it arrived at the target's AOI before the agent's hand; conversely, it was considered non-predictive (negative score) if the child's gaze arrived at the target's AOI after the agent's hand [36].

We used a clearview fixation filter for the eye-tracking data. Fixation was defined as gaze recorded within a 50-pixel diameter for a minimum of 200 ms, and this criterion was applied to the raw eye-tracking data to determine the duration of any fixation.

To assess the infants' action anticipation for the two conditions (human and robot), we carried out two *t*-tests to compare the anticipatory gaze with the reaching time. Furthermore, to assess difference in the anticipatory gaze between human and robot, we carried out a *t*-test comparing both conditions. In case of a null result, we also included BF_{01}, indicating the Bayes factor in favor of the H0 over H1.

To assess infants' general attentional pattern to the video stimuli, total fixation duration was entered in three repeated measures General Linear Models (GLMs) analysis: (1) the first GLM aimed to analyse children's general attention to the face in the three sequences with 3 levels of Sequence (*Sequence*-1, *Sequence*-2, *Sequence*-3) that have the same duration (2 s) as within-subjects factors, and and 2 levels of *Agent*(Human, Robot) as the between-subjects factor; (2) the second GLM aimed to analyse the general attention to the face and hand in Sequence-2 with 2 levels of *AOIs* (Face, Hand) as within subjects factors, and 2 levels of *Agent*(Human, Robot) as within-subjects factors, and 2 levels of agency (Human, Robot) as the between-subjects factor; (3) the third GLM aimed to analyse the general attention to the face and target in Sequence-3 with 2 levels of *AOIs* (Face, Target) as within-subjects factors, and 2 levels of *Agent*(Human, Robot) as the between-subjects factor. The data were normally distributed, as assessed by inspection of the Q–Q Plots and Skewness (range = 0.43–0.62). The Greenhouse–Geisser correction was used for violations of Mauchly's Test of Sphericity ($p < 0.05$). F-tests post-hoc comparisons were Bonferroni corrected.

3 Results

3.1 Action Prediction

To evaluate the infants action prediction, we conducted a *t*-test comparing the anticipatory gaze and the reaching time. The *t*-test comparing the anticipatory gaze and reaching time (time zero) showed a significant difference both in Human condition and Robot condition (anticipation time > reaching time; Human: Mean Anticipation Time = 0.73; $SD = 0.68$; $t(13) = 9.57, p < 0.001, d = 2.56$; Robot: Mean Anticipation Time = 0.73; $SD = 0.59$; $t(11) = 6.1, p < 0.001, d = 1.75$).

Furthermore, we conducted a *t*-test to compare action prediction between Human and Robot conditions. The *t*-test comparing the anticipation time between the two agents (Human and Robot) did not reveal a difference ($p > 0.05$). The results are summarized in Fig. 3. To confirm that infants'

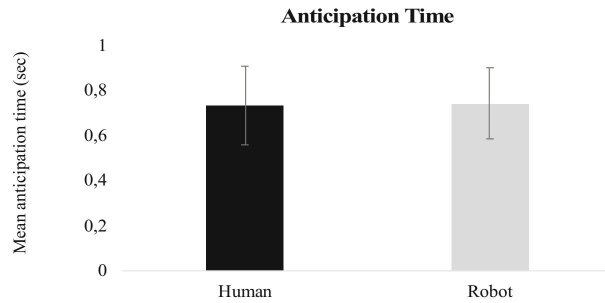

Fig. 3 Infants' action prediction, i.e. the time of gaze arrival in the target' AOI while the agent moves the arm to touch the target (Sequence-3). Positive scores indicate that the infants predicted the action. The graph shows the mean scores (sec) of the infants' anticipatory gaze as a function of the agent (Human, Robot). The bars represent the standard error of the mean

anticipatory gaze (i.e., anticipation of action) does not differ between human and robot, we conducted a *t*-test using a Bayesian approach, $BF_{01} = 3.41$, Median = − 0.187, 95% CI: [− 0.688, − 0.008].

3.2 Total Fixation Duration: Face Across the Three Sequences

To assess infants' general attentional pattern to face, we conducted a repeated measures GLM for the total fixation duration across the three sequences. The results showed a main effect of *Agent*, $F(1, 27) = 8.78$, $p < 0.01$, *partial*-$\eta^2 = 0.25$, indicating that, independent of sequence, infants exhibited a greater fixation time on the face for the Robot (Mean Total Fixation = 1.47, $SD = 0.13$) compared to the Human (Mean Total Fixation = 0.92, $SD = 0.12$), Mdiff = 0.54, SE = 0.18, $p < 0.01$ (Fig. 4A). Additionally, the results revealed a main effect of *Sequence*, $F(2, 54) = 131.157$, $p < 0.001$, *partial*-$\eta 2 = 0.83$, indicating that, independent of agent, infants exhibited a greater fixation time on the face in Sequence-1 (Mean Total Fixation = 1.92, $SD = 0.14$), compared to both Sequence-2 (Mean Total Fixation = 0.92, $SD = 0.12$), Mdiff = 0.802, SE = 0.07, $p < 0.001$, and Sequence-3 (Mean Total Fixation = 0.48, $SD = 0.06$), Mdiff = 1.47, SE = 0.11, p < 0.001 (Fig. 4B). Furthermore, infants' fixation time on the face was lower to the Sequence- 3 compared to both Sequence-1, Mdiff = − 0.66, SE = 0.07, $p < 0.001$ and Sequence-2, Mdiff = − 1.47, SE = 0.12, $p < 0.001$ (Fig. 4B).

3.3 Total Fixation Duration: Face and Hand in Sequences 2

To assess infants' looking behavior before action prediction, we conducted a repeated measures GLM for the total fixation duration in Sequences-2 to Face and Hand. The results showed a main effect of *AOIs*, $F(1, 27) = 28.06$, $p < 0.01$, *partial*-$\eta^2 = 0.51$, indicating that, independent of agent,

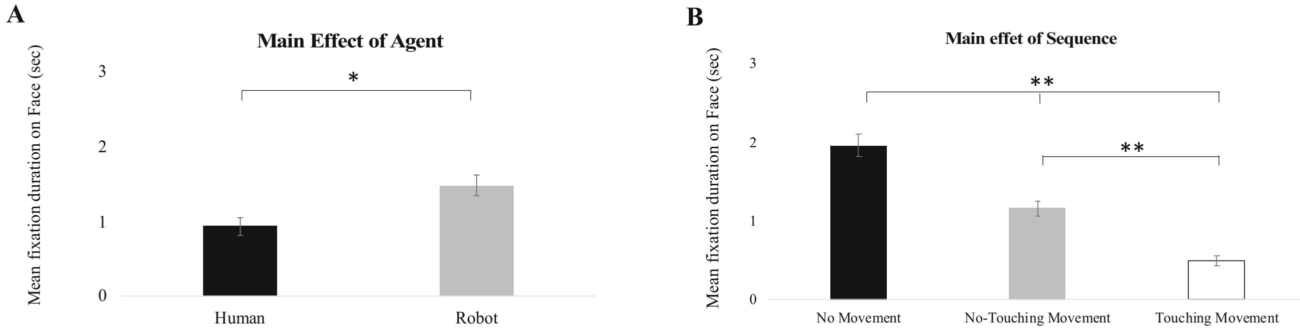

Fig. 4 **A–B** Attention paid (total fixation duration) by the infants to the face of the human and the robot across the three phases of the video: In the **A** the graph shows the total fixation duration mean scores (sec) as a function of the agent (Human, Robot); in the **B** the graph shows the total fixation duration mean scores (sec) as a function of the sequence (Sequence-1, Sequence-2, Sequence-3). The bars represent the standard error of the mean. *Indicates significant differences = p < 0.01, and ** = p < 0.001

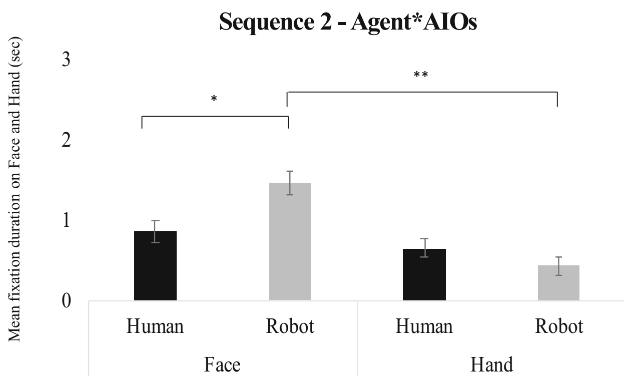

Fig. 5 Attention paid (total fixation duration) by the infants to the face and hand of the human and the robot in the Sequence-2 in which the agent watches the hand lift slightly from the table. The graph shows the mean scores (sec) as a function of agent (Human, Robot) and area of interest (Face, Hand). The bars represent the standard error of the mean. *Indicates significant differences = p < 0.01, and ** = p < 0.001

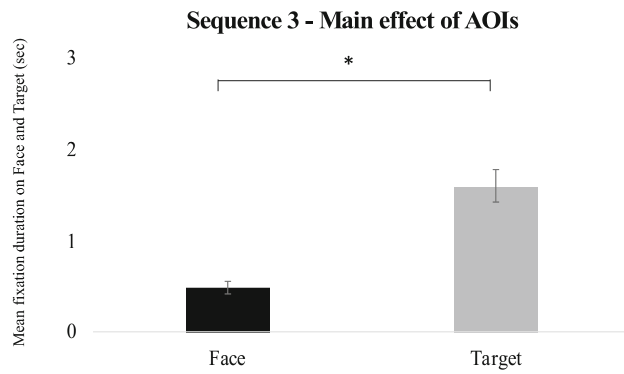

Fig. 6 Attention paid (total fixation duration) by the infants to the face of the human and robot and to the target in the Sequence-3 in which the agent moved the arm reaching the target. The graph shows the mean scores (sec) for each area of interest (Face, Target) across agent. The bars represent the standard error of the mean. *Indicates significant differences = p < 0.001

infants exhibited a greater fixation time on the face (Mean Total Fixation = 1.15, *SD* = 0.11) compared to the hand (Mean Total Fixation = 0.53, *SD* = 0.12), *M*diff = 0.62, SE = 0.12, $p < 0.001$. Additionally, the results revealed a significant interaction between *Agent*AOIs*, $F(1, 27) = 12.66$, $p < 0.005$, *partial-η2* = 0.32, indicating that infants exhibited a greater fixation time on the face for the Robot condition (Mean Total Fixation = 1.45, *SD* = 0.14) compared to the Human condition (Mean Total Fixation = 0.85, *SD* = 0.13), *M*diff = 0.61, SE = 0.19, $p < 0.01$, and that for the Robot condition the face is fixated longer than hand (Mean Total Fixation = 0.42, *SD* = 0.12), *M*diff = 1.03, SE = 0.17, $p < 0.001$. This interaction is plotted in Fig. 5.

3.4 Total Fixation Duration: Face and Target in Sequences 3

To assess the looking behavior during watching reaching action, we conducted a repeated measures GLM for the total fixation duration in Sequences-3 to Face and Target. The results showed a main effect of *AOIs*, $F(1, 27) = 34.23$, $p < 0.01$, *partial-*$\eta^2 = 0.56$, indicating that, independent of agent, infants exhibited a greater fixation time on the target (Mean Total Fixation = 1.6, *SD* = 0.18) compared to the face (Mean Total Fixation = 0.48, *SD* = 0.06), *M*diff = 1.11, SE = 0.19, $p < 0.001$. This main effect is plotted in Fig. 6.

4 Discussion

The present study investigated the ability of 17-month-old infants to anticipate an action when performed by a human and a humanoid robot. For this purpose, the anticipatory gaze behavior of infants was measured while watching videos in which both a human and a humanoid robot (i.e., Robovie) moved their arm to touch a toy placed on a table in front of them. In general, results on action anticipation showed no differences in infants' anticipatory gaze on the target object between the two agents. Nevertheless, infants showed greater general attention to the robot's face than human's face. When reaching the target (i.e., Sequence-3), the infants rightly focused their attention on the target independent of the agent.

More specifically, the results regarding anticipatory gaze for the Human condition confirm our first hypothesis, i.e., that infants at 17 months anticipate the action of the human agent. This result is in line with the literature showing that infants within the first year of life anticipate simple actions (e.g., reaching a toy) performed by familiar agents (i.e., humans) [9, 12–21, 36, 38]. Furthermore, the data on anticipatory gaze for the Robot condition confirm our second hypothesis, namely that infants at 17 months anticipate the action of a humanoid robot. This finding, on the one hand, confirm the idea that, from the middle of the second year of life, infants anticipate actions of non-human agents [38] and, on the other hand, enrich the findings on the ability of 18-months-old infants to predict action of mechanical arms [38]. As matter of fact, while our findings suggest a similar phenomenon as that found for the mechanical arm with 17-month-olds when the action is performed by a humanoid robot, they do not confirm our third hypothesis, namely that 17-month-olds anticipate human action faster than humanoid robot action. This result show that infants from 17 months of age can predict the actions of anthropomorphized non-human agents (i.e., humanoid robot) as accurately as for a human agent.

Taken together, these results could be explained by the direct-matching hypothesis [50, 51]. This hypothesis claims that the action of another is understood when its observation causes the motor system of the observer to resonate [51]. This mechanism has been found from 6-month-old infants for simple actions performed by human agents (familiar agent) and in adults also with non-human agents (unfamiliar agent) [36]. The act of reaching a toy—simple and familiar action—is largely encoded in the 17-month-old child's motor repertoire [20, 52–54] and the agent is familiar (for a review see [9]). These two components together—simple action and familiar agent—would prompt the child to anticipate the human's action from early infancy [9]. However, in our study we found that, although the humanoid robot was not a familiar agent for the infant, it was actually able to elicit the same processes evoked by the human agent. This apparent discrepancy is partially explained by infant research that suggests that up to the first year of life, agent familiarity is a crucial component for action prediction [9], whereas infants at 18 months anticipate simple actions of unfamiliar non-human agents (mechanical arms; [38]). Our study enriches the literature by showing that motor anticipation toward the action of a humanoid robot is similar to that of a human, a condition that was lacking, for example, in the study by Adam et al. [38].

From a developmental perspective, these results seem to witness a decrease in the salience of the familiarity with the agent performing the action compared to familiarity with the action itself. The main result of our study—i.e., similar action anticipation times for the human and the humanoid robot—, considering direct matching mechanism and developmental findings, could then be explained in at least twofold manner. On the one hand, the anthropomorphization of the humanoid robot could increase the degree of familiarity of the non-human agent for infants, mitigating their unfamiliarity with the agent in their motor repertoire. The effect of anthropomorphization on children's responses has been extensively studied with humanoid robots, showing that the robot human resemblance positively influences children's responses towards the robot [7, 55–57] (keeping in mind that the extreme anthropomorphization of the robot could lead to the Uncanny Valley Effect, [58]). On the other hand, the extensive familiarity of 17-month-olds with the simple action of reaching a toy may have tempered the effect of unfamiliarity with the agent. This interpretation is in line with the literature indicating that with experience children become increasingly accurate to anticipate even complex actions with human [13, 36, 59, 60].

A second body of data from the present study, relate to the attention paid by infants to the human and the humanoid robot. While no differences were found in terms of anticipation of action between human and robot, the data showed different attentional patterns for the two agents with respect to the AOIs (Face, Hand, and Target) and the sequence of the video stimuli (Sequence-1, Sequence-2, Sequence-3). In general, independent of the video sequence infants focused more on the robot's face than on the human's face and, at the same time, regardless of the agent the infants' attention to the face decreased between sequences. In particular, in Sequence-3 attention was greater to the target than to the face, showing the effectiveness of the stimuli in shifting the infants' attention to the target.

Regarding data on the face, a wide body of evidenced demonstrated that the face draws particular attention from the early infancy in social interactions [61–64]. As early as 6 months of age, the face of humanoid robots particularly captures the attention of infants [55]. Matsuda et al. [55] showed that infants between 6 and 14 months of age are more attuned to the face of a humanoid robot (marked by some human resemblance) than to the face of both a human and an android (characterized by extreme resemblance to a human), attributing this difference to a novelty effect due to the humanoid robot. Our results extend to 17-month-old infants the greater amount of time devoted to the face of a humanoid robot compared with a human face. Moreover, as hypothesized by Matsuda et al. [55], the infants in our study looked at the robot's face longer than the human's face because the robot represents a novel stimulus for them (Robovie2 is a robot that is not commercialized but used as a research platform). At the same time, this result preliminarily supports our interpretation of the effect of robot anthropomorphization as a mediator of agent unfamiliarity in action prediction in the robot condition. Indeed, it is plausible that infants spent more time observing the robot's face to compensate their unfamiliarity with the robotic agent. This is speculation should be further investigated in future studies.

5 Concluding Remarks, Limitations of the Study and Future Directions

The present study allowed to highlight the ability to predict action goals of a human and a humanoid robot and the underlying attentional patterns in 17-month-old infants. Specifically, the results showed that infants can anticipate the goal-directed action of both human and robot, suggesting that 17-month-olds anticipate simple actions of non-human anthropomorphic agents. Furthermore, the results indicated that the robot face is an important component of 17-month-old infants' information processing. Overall, these results provide new insights for both the field of developmental psychology, showing that anthropomorphization of the agent performing the action can mitigate the effect of unfamiliarity with the non-human agent in the action prediction, as well as for the field of robotics, highlighting that humanoid robots due to their human-like characteristics can foster the activation of social mechanisms already in the first months of life.

The study is not without some limitations that could be addressed in the future. A first aspect relates to the generalization of our results to actions that are complex, ambiguous and unfamiliar to infants. Indeed, our results are based on a simple, unambiguous and known action (i.e., reaching a toy) performed by a familiar (human) and unfamiliar (humanoid robot) agent. Future studies could use a different set of actions to disentangle the effect between familiarity of the action and unfamiliarity of the agent. With respect to agent familiarity, a second limitation is the generalization to all types of humanoid robots. Several studies on humanoid robots show that different levels of physical anthropomorphization of the robot have different effects on the attribution of mental qualities and psychological characteristics from the age of 3 years [4, 7]. This proclivity of infants to be influenced by the anthropomorphization of humanoid robots could have a general effect at earlier ages and, specifically, on action prediction. Therefore, it would be important for future studies to use different types of robots (from the more mechanical to the more anthropomorphic) to understand if infants' ability to predict the action of humanoid robots is independent of the degree of robot anthropomorphization. Finally, a third aspect concerns the age of the infants. In the present study, we showed the ability of 17-month-old infants to predict the action of a humanoid robot, however, we do not actually know if this ability occurs before this age. Future studies could extend the analysis of action prediction of humanoid robots to different ages suggesting possible developmental patterns in the anticipation of a robotic agent's action.

Author contributions F.M. and M.I. conceived the experiment. M.I. conducted the experiment. F.M., M.I. and C.D.D. analyzed the results. All authors contributed to the writing of the manuscript.

Funding Open access funding provided by Università Cattolica del Sacro Cuore within the CRUI-CARE Agreement. This research was funded by the Japanese Society for the Promotion of Science, Programme Grant # 16H01880, 16H06301, 15H01846, 25245067. In addition, we would like to thank the entire Baby Science Centre of Doshisha University and Kyoto University, especially Fumina Sano and Nanami Toya, for data collection. Also, Università Cattolica del Sacro Cuore contributed to the funding of this research project and its publication (research funding action D.3.1).

Data availability Data available on request from the authors.

Declarations

Competing interests The author(s) declare no competing interests.

References

1. Fitter NT, Funke R, Pulido JC et al (2019) Socially assistive infant-robot interaction: using robots to encourage infant leg-motion training. IEEE Robot Autom Mag 26:12–23. https://doi.org/10.1109/MRA.2019.2905644
2. Marchetti A, Di Dio C, Manzi F, Massaro D (2022) Robotics in clinical and developmental psychology. In: Reference Module in neuroscience and biobehavioral psychology. Elsevier, p B9780128186978000000
3. Marchetti A, Manzi F, Itakura S, Massaro D (2018) Theory of mind and humanoid robots from a lifespan perspective. Z Für Psychol 226:98–109. https://doi.org/10.1027/2151-2604/a000326

4. 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
5. Di Dio C, Manzi F, Peretti G et al (2020) Come i bambini pensano alla mente del robot. Il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. Sist Intelligenti 41–56. https://doi.org/10.1422/96279
6. 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 Soc Robot 12:1045–1059. https://doi.org/10.1007/s12369-019-00528-9
7. 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
8. Hunnius S, Bekkering H (2014) What are you doing? How active and observational experience shape infants' action understanding. Philos Trans R Soc B Biol Sci 369:20130490. https://doi.org/10.1098/rstb.2013.0490
9. Elsner B, Adam M (2021) Infants' goal prediction for simple action events: the role of experience and agency cues. Top Cogn Sci 13:45–62. https://doi.org/10.1111/tops.12494
10. Woodward A (1998) Infants selectively encode the goal object of an actor's reach. Cognition 69:1–34. https://doi.org/10.1016/S0010-0277(98)00058-4
11. Meltzoff AN (1995) Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. Dev Psychol 31:838–850. https://doi.org/10.1037/0012-1649.31.5.838
12. Gredebäck G, Falck-Ytter T (2015) Eye movements during action observation. Perspect Psychol Sci 10:591–598. https://doi.org/10.1177/1745691615589103
13. Ambrosini E, Reddy V, de Looper A et al (2013) Looking ahead: anticipatory gaze and motor ability in infancy. PLoS ONE 8:e67916. https://doi.org/10.1371/journal.pone.0067916
14. Brandone AC, Horwitz SR, Aslin RN, Wellman HM (2014) Infants' goal anticipation during failed and successful reaching actions. Dev Sci 17:23–34. https://doi.org/10.1111/desc.12095
15. Cannon EN, Woodward AL, Gredebäck G et al (2012) Action production influences 12-month-old infants' attention to others' actions: action production and anticipation. Dev Sci 15:35–42. https://doi.org/10.1111/j.1467-7687.2011.01095.x
16. Cannon EN, Woodward AL (2012) Infants generate goal-based action predictions: infants generate goal-based predictions. Dev Sci 15:292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x
17. Falck-Ytter T, Gredebäck G, von Hofsten C (2006) Infants predict other people's action goals. Nat Neurosci 9:878–879. https://doi.org/10.1038/nn1729
18. Gredebäck G, Fikke L, Melinder A (2010) The development of joint visual attention: a longitudinal study of gaze following during interactions with mothers and strangers: the development of joint visual attention. Dev Sci 13:839–848. https://doi.org/10.1111/j.1467-7687.2009.00945.x
19. Hofsten C, Dahlström E, Fredriksson Y (2005) 12-Month-old infants' perception of attention direction in static video images. Infancy 8:217–231. https://doi.org/10.1207/s15327078in0803_2
20. Green D, Li Q, Lockman JJ, Gredebäck G (2016) Culture influences action understanding in infancy: prediction of actions performed with chopsticks and spoons in Chinese and Swedish infants. Child Dev 87:736–746. https://doi.org/10.1111/cdev.12500
21. Stapel JC, Hunnius S, Meyer M, Bekkering H (2016) Motor system contribution to action prediction: temporal accuracy depends on motor experience. Cognition 148:71–78. https://doi.org/10.1016/j.cognition.2015.12.007
22. Király I, Jovanovic B, Prinz W et al (2003) The early origins of goal attribution in infancy. Conscious Cogn 12:752–769. https://doi.org/10.1016/S1053-8100(03)00084-9
23. Woodward AL, Sommerville JA (2000) Twelve-month-old infants interpret action in context. Psychol Sci 11:73–77. https://doi.org/10.1111/1467-9280.00218
24. Hunnius S, Bekkering H (2010) The early development of object knowledge: a study of infants' visual anticipations during action observation. Dev Psychol 46:446–454. https://doi.org/10.1037/a0016543
25. Hunnius S, Bekkering H, Cillessen AHN (2009) The association between intention understanding and peer cooperation in toddlers. Int J Dev Sci 3:368–388. https://doi.org/10.3233/DEV-2009-3404
26. Kochukhova O, Gredebäck G (2010) Preverbal infants anticipate that food will be brought to the mouth: an eye tracking study of manual feeding and flying spoons: infants comprehension of feeding. Child Dev 81:1729–1738. https://doi.org/10.1111/j.1467-8624.2010.01506.x
27. Melzer A, Prinz W, Daum MM (2012) Production and perception of contralateral reaching: a close link by 12 months of age. Infant Behav Dev 35:570–579. https://doi.org/10.1016/j.infbeh.2012.05.003
28. Bekkering H, Wohlschläger A, Gattis M (2000) Imitation of gestures in children is goal-directed. Q J Exp Psychol Sect A 53:153–164. https://doi.org/10.1080/713755872
29. Gergely G, Bekkering H, Király I (2002) Rational imitation in preverbal infants. Nature 415:755–755. https://doi.org/10.1038/415755a
30. Marshall PJ, Meltzoff AN (2011) Neural mirroring systems: exploring the EEG mu rhythm in human infancy. Dev Cogn Neurosci 1:110–123. https://doi.org/10.1016/j.dcn.2010.09.001
31. Southgate V, Begus K, Lloyd-Fox S et al (2014) Goal representation in the infant brain. Neuroimage 85:294–301. https://doi.org/10.1016/j.neuroimage.2013.08.043
32. Deubel H, Schneider WX (1996) Saccade target selection and object recognition: evidence for a common attentional mechanism. Vis Res 36:1827–1837. https://doi.org/10.1016/0042-6989(95)00294-4
33. Daum MM, Gredebäck G (2011) The development of grasping comprehension in infancy: covert shifts of attention caused by referential actions. Exp Brain Res 208:297–307. https://doi.org/10.1007/s00221-010-2479-9
34. Daum MM, Wronski C, Harms A, Gredebäck G (2016) Action perception in infancy: the plasticity of 7-month-olds' attention to grasping actions. Exp Brain Res 234:2465–2478. https://doi.org/10.1007/s00221-016-4651-3
35. Adam M, Reitenbach I, Papenmeier F et al (2016) Goal saliency boosts infants' action prediction for human manual actions, but not for mechanical claws. Infant Behav Dev 44:29–37. https://doi.org/10.1016/j.infbeh.2016.05.001
36. Kanakogi Y, Itakura S (2011) Developmental correspondence between action prediction and motor ability in early infancy. Nat Commun 2:341
37. Adam M, Reitenbach I, Elsner B (2017) Agency cues and 11-month-olds' and adults' anticipation of action goals. Cogn Dev 43:37–48. https://doi.org/10.1016/j.cogdev.2017.02.008
38. Adam M, Gumbsch C, Butz MV, Elsner B (2021) The impact of action effects on infants' predictive gaze shifts for a non-human grasping action at 7, 11, and 18 months. Front Psychol 12:695550. https://doi.org/10.3389/fpsyg.2021.695550
39. Manzi F, Di Dio C, Itakura S et al (2020) Moral evaluation of human and robot interactions in Japanese pre-schoolers. In: CEUR workshop proceedings. Germany, pp 20–27
40. 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

41. Okumura Y, Kanakogi Y, Kanda T et al (2013) Infants understand the referential nature of human gaze but not robot gaze. J Exp Child Psychol 116:86–95. https://doi.org/10.1016/j.jecp.2013.02.007
42. Okumura Y, Kanakogi Y, Kanda T et al (2013) The power of human gaze on infant learning. Cognition 128:127–133. https://doi.org/10.1016/j.cognition.2013.03.011
43. Okumura Y, Kanakogi Y, Kanda T et al (2013) Can infants use robot gaze for object learning?: The effect of verbalization. Interact Stud Soc Behav Commun Biol Artif Syst 14:351–365. https://doi.org/10.1075/is.14.3.03oku
44. Meltzoff AN, Brooks R, Shon AP, Rao RPN (2010) “Social” robots are psychological agents for infants: a test of gaze following. Neural Netw 23:966–972. https://doi.org/10.1016/j.neunet.2010.09.005
45. 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
46. Sommer K, Redshaw J, Slaughter V et al (2021) The early ontogeny of infants’ imitation of on screen humans and robots. Infant Behav Dev 64:101614. https://doi.org/10.1016/j.infbeh.2021.101614
47. Flanagan JR, Johansson RS (2003) Action plans used in action observation. Nature 424:769–771. https://doi.org/10.1038/nature01861
48. Koch B, Stapel J (2019) The role of head and hand movements for infants’ predictions of others’ actions. Psychol Res 83:1269–1280. https://doi.org/10.1007/s00426-017-0939-6
49. Fausey CM, Jayaraman S, Smith LB (2016) From faces to hands: changing visual input in the first two years. Cognition 152:101–107. https://doi.org/10.1016/j.cognition.2016.03.005
50. Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230
51. Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci 2:661–670. https://doi.org/10.1038/35090060
52. Nyström P, Ljunghammar T, Rosander K, von Hofsten C (2011) Using mu rhythm desynchronization to measure mirror neuron activity in infants: measuring mirror neuron activity in infants. Dev Sci 14:327–335. https://doi.org/10.1111/j.1467-7687.2010.00979.x
53. Southgate V, Johnson MH, Osborne T, Csibra G (2009) Predictive motor activation during action observation in human infants. Biol Lett 5:769–772. https://doi.org/10.1098/rsbl.2009.0474
54. Southgate V, Johnson MH, Karoui IE, Csibra G (2010) Motor system activation reveals infants’ on-line prediction of others’ goals. Psychol Sci 21:355–359. https://doi.org/10.1177/0956797610362058
55. Matsuda G, Ishiguro H, Hiraki K (2015) Infant discrimination of humanoid robots. Front Psychol. https://doi.org/10.3389/fpsyg.2015.01397
56. Tung FW (2016) Child perception of humanoid robot appearance and behavior. Int J Hum Comput Interact
57. Woods S (2006) Exploring the design space of robots: children’s perspectives. Interact Comput 18:1390–1418. https://doi.org/10.1016/j.intcom.2006.05.001
58. Mori M. The uncanny valley. Energy 7:33–35
59. Henderson AME, Wang Y, Matz LE, Woodward AL (2013) Active experience shapes 10-month-old infants’ understanding of collaborative goals: experience and collaboration. Infancy 18:10–39. https://doi.org/10.1111/j.1532-7078.2012.00126.x
60. Needham A, Libertus K (2011) Embodiment in early development. Wiley Interdiscip Rev Cogn Sci 2:117–123. https://doi.org/10.1002/wcs.109
61. Farroni T, Csibra G, Simion F, Johnson MH (2002) Eye contact detection in humans from birth. Proc Natl Acad Sci 99:9602–9605. https://doi.org/10.1073/pnas.152159999
62. Gliga T, Elsabbagh M, Andravizou A, Johnson M (2009) Faces attract infants’ attention in complex displays. Infancy 14:550–562. https://doi.org/10.1080/15250000903144199
63. Simion F, Macchi Cassia V, Turati C, Valenza E (2001) The origins of face perception: specific versus non-specific mechanisms. Infant Child Dev 10:59–65. https://doi.org/10.1002/icd.247
64. Johnson MH, Dziurawiec S, Ellis H, Morton J (1991) Newborns’ preferential tracking of face-like stimuli and its subsequent decline. Cognition 40:1–19. https://doi.org/10.1016/0010-0277(91)90045-6