

On the Imitation of Goal Directed Movements of a Humanoid Robot

Yunqing Bao¹ · Raymond H. Cuijpers¹

Accepted: 3 June 2017 / Published online: 23 June 2017
© The Author(s) 2017. This article is an open access publication

Abstract Interacting with a social robot should give people a better understanding of the robot’s actions and intentions. In terms of human–human interaction (HHI), people can interpret actions of others in an effortless way. However, it is still unclear whether people can do the same with humanoid robots. Imitation (of the robot’s actions) provides us with an intuitive means for solving this puzzle, because it is closely related to interpreting another’s actions. In the study of human imitation, the theory of goal-directed imitation holds that the imitator tends to imitate the action goals whenever the action goals are salient, otherwise people tend to imitate the action means. We investigated this theory for human robot interaction by manipulating the presence and absence of the goal object when people imitate the robot’s pointing gestures. The results showed that the presence of a goal object reduces people’s goal errors. Moreover, we found that most people tend to match their action means rather than the goals of the robot’s action. To ensure that participants considered the robot as a social agent, we designed a natural interaction task where the turn-taking cue was included: we let the robot look at the human after completing the pointing gesture at varying latencies. As expected, we found that the earlier the robot gazes at people, the shorter the reaction time of people was to start to imitate. Our results show that people are responsive to a robot’s social gaze cues, and that they are responsive to the action goals of robots, although not as much as in HHI.

Keywords Theory of mind · The GOADI theory · Human–robot Interaction · Turn-taking · Eye-gaze

1 Introduction

Humanoid robots are expected to work together with human-beings in everyday environments in the future. Under this circumstance, they need to be able to interact with humans in a safe and natural way. One important aspect for a successful human–robot interaction (HRI) lies in whether people consider the robot as a potential social agent [1,2]. Interacting with a social robot should give people a better understanding of the robot’s actions and intentions. Therefore, as emphasized by Scassellati [3] and other researchers [4–6], it is crucial to focus on how people naturally adopt an ‘intentional stance’ and interpret the behaviour of the robot as if it possesses goals, intentions, and beliefs.

In terms of human–human interaction (HHI), humans demonstrate a remarkable ability to interpret the actions and intentions of others in a seemingly effortless way [7]. For instance, when you see somebody stretching out his hand toward a cup, you do not just see a hand movement trajectory, but you also infer the goal of reaching that cup, i.e. drinking. It stands to reason that, although robots may be humanoid, their movement kinematics still differ substantially from those of humans. However, Gazzola et al. [8] found that the mirror neuron system (MNS) activates both for robotic and humanoid movements. They reasoned that the mirror neuron system might be responding to the action goal. Yet, if a humanoid robot is stretching out its hand toward a cup, it is not immediately obvious that an observer still understands its intention. Surely one will not assume that the robot intends to drink? Whether or not people

✉ Yunqing Bao
Y.Bao@tue.nl

Raymond H. Cuijpers
R.H.Cuijpers@tue.nl

¹ Human Technology Interaction, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

infer goals from observed robot actions depends on whether the robot is treated as an intentional agent or as a mindless machine. In other words, it depends on the observer's mental model of the robot. In HHI, human observers form mental representations of the observed actions, which allows them to predict and interpret observed behaviour. It is still unclear to what extent similar mental representations are formed for humanoid robots and whether it allows predicting and interpreting of the robot's behaviour [9].

There are two main views on how people interpret observed actions from others. The theory of mind (ToM) holds that people interpret observed behaviour from other people according to a generic model of human behaviour [10]. Conversely, simulation theory (ST) claims that people simulate the observed actions using their own action system in order to make predictions from which action goals are inferred [11]. A potential neural substrate for ST are the so-called 'mirror neurons', which fire both when observing and when performing an action [8, 12, 13]. It was shown that mirror neurons are involved in action understanding [14], and imitation [15, 16]. For example, it was shown that there is a decrease in the latencies of actions by observing matching actions. This is taken as evidence in support of ST, as the mirror neurons appear to simulate the brain state of another [17]. Being able to internally simulate observed actions in order to infer associated action goals is believed to provide a common ground between interacting agents. For example, in joint action reaching a common action goal may require complementary actions rather than imitative actions [18], so only the action goals can be "imitated". Recognising the goals of goal-directed actions does not necessarily mean that the observed actor is perceived as an intentional agent. This would require mental representations of behaviour of intentional agents, similar to what is proposed by ToM. Current views include the possibility that both ST and ToM are at work in humans [19].

For HRI, imitating goals is important, because robots typically have different bodies which further complicates the question of 'what to imitate?' (e.g. [20]). Inferring action goals allows "imitation" without the need to use the exact same means [20, 21]. In [22] a dynamic framework is presented that enables robots to infer goals from observed actions. There is also evidence that humans infer action goals of robots. It was found that mirror neuron activity is very similar when a human observer sees a human hand or an industrial robot hand performing similar actions. Moreover, the mirror regions of the brain responding to the sight of industrial robotic actions responded more during the observation of goal-directed actions than similar movements not directed at goals. Oberman et al. [23] found evidence of mirror neuron system activation in interaction with humanoid robots using EEG.

1.1 Theory of Goal-Directed Imitation

Bekkering et al. [24] found that people prefer imitating the goal of a goal-directed movement over its means: in their study, a human presenter was sitting face-to-face with the follower, and the presenter's movement was either contra-lateral (using the right hand to touch the left ear or vice versa) or ipsi-lateral (using the right hand to touch the right ear or the left hand to touch the left ear). The followers were asked to imitate the presenter's movement as if looking in a mirror. It turned out that the followers made many ipsi-lateral movements towards the correct ear even when imitating contra-lateral movements. Thus, the correct movement trajectory was 'sacrificed' to achieve the correct target in a goal-directed movement. This form of 'imitation' is also referred to as *emulation*. Lyons [25] further refines imitating by noting that tasks are typically composed of a hierarchy of goals and subgoals, which may be imitated or emulated depending on the observer's judgement. Wohlschläger et al. [16, 26] studied this object-oriented imitation pattern by manipulating the presence and absence of the goal object in order to see how goals are organized in imitation. Note that in the touch ear experiment [24], the goal objects (ears) were always present. Therefore they replaced the 'hand touch one of two ears' with the 'hand touch one of two dots on the table' task [16]. It has been shown that when the dots are present, dot selection becomes the highest goal of imitation. However, when the dots are absent, the hand and/or movement trajectory are the highest goal of imitation. Moreover, they also found systematic error patterns in imitation similar to the touching ear experiment: when the experimenter uses contra-lateral movement to touch a dot, people tend to imitate with touching the same dot as the experiment does; nevertheless, they use the ipsi-lateral movement path to imitate.

On the basis of these systematic error patterns, the theory of goal-directed imitation (GOADI) was proposed [26]. GOADI holds that imitation is based on the identification of action goals and on their organization into hierarchical structures. The action goals may be objects, effector (the hand), positions in space, or movement trajectories [27]. When a goal object is present, target object selection becomes the highest goal of imitation. However, when a goal object is absent, people tend to use the correct hand and/or movement trajectory as the highest goal of imitation. In other words, the selected goals are hierarchically ordered in this way: the ends (the selection of the present goal object) are more important than the means (the effectors or the movement path).

Currently, the GOADI theory was only verified in HHI, and it has not been shown to apply to HRI. Verifying the GOADI theory in HRI helps us understand better how people interpret robot behaviour. In particular, if people tend to imitate a robot's action goals, we may expect that people use

their own action system to infer the goals of observed movements. This is important for knowing how to design robot behaviour and to understand which behaviours are considered “natural”.

1.2 Social Gaze Cues

Action observation and goal-inference are needed to establish a common ground for joint attention [18], and joint attention is thought to underlie many types of social interaction [28]. But being able to extract the goal of a robot’s movement is hardly sufficient to claim that people treat the robot as a ‘social agent’ rather than a machine. To attribute mental states to (artificial) agents requires an intentional stance [6]. ToM can be regarded as the fundamental link for achieving successful social interaction (in HHI), which ST cannot fully explain [3, 19]. Therefore, it is important to look into what kind of social cues could influence people’s mental model of the robot, and thus, perceive the robot as a ‘social agent’.

According to Williams et al. [29], drawing people’s attention to objects within perceptual vicinity is a critically important ‘social skill’ for a robot. Moreover, attention is necessary for activation of an action representation [30]. Williams and colleagues [29] implemented the joint-attention mechanism on a navigating robot and proved its effectiveness experimentally. They demonstrated that people can interpret robot pointing behaviour in a similar way as human pointing behaviour. Moreover, people tend to take gaze cues (the head trajectory) into consideration when they determine the robot’s pointing target.

Gaze and eye contact are well-known cues for many social interactions including turn-taking [28, 31]. For example, gaze following of an avatar supports prediction of its movements [32] and people engage faster with a robot employing gaze cues [33]. Turn-taking occurs when a speaker expects some reaction from the listener [34]. Typically, a speaker will shift his/her gaze to the listener to indicate the end of a turn. According to Ito and Tani [35], turn-taking is considered to be a prerequisite for joint attention because the function of turn-taking is to initiate the joint attention. As joint attention requires mutual awareness of the companion’s attention, turn-taking regulates switching initiative in interactions among agents [36]. Humans typically ‘know’ when to start and stop their turns in the social interactions in HHI, based on various factors like the (semantic) context and nonverbal feedback. Gaze is one of the most frequently studied social cues in HHI and has been shown to provide useful information that regulates a given social interaction and facilitates accomplishment of a task or a goal [2]. Jokinen [37] observed that the listener’s responses were quicker if there was a mutual gaze and they were delayed if the previous utterance terminated without a speaker gaze. Likewise, Gu and Badler [38] pointed out that social gaze cues provide

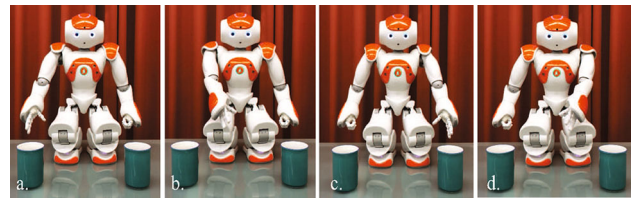


Fig. 1 A front view of Nao’s hand movements

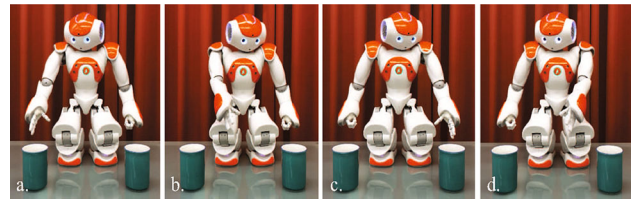


Fig. 2 A front view of the combination of Nao’s hand and gaze

turn-taking signals to regulate the flow in a dyadic conversation.

1.3 The Current Study

In the current study, we looked into whether the GOADI theory, which was initially tested in HHI, also applied to HRI. We verify whether participants can imitate the goal of the robot’s action rather than the exact means of the robot’s action. In the initial GOADI experiment the human demonstrator used either hands to touch one of two dots. Since it was relatively difficult to implement this behaviour on our humanoid robot, we simplified this to hand pointing and gaze pointing towards two cups (see Figs. 1, 2). Similar to the original GOADI experiments the goal object could be present or absent. The gaze pointing behaviour supported joint attention, but it also enabled a turn-taking cue by shifting the gaze back to a person. In order to investigate whether the turn-taking cues would affect the timing of imitation in HRI, we manipulated the gaze timing at which the robot looked back at participants with respect to its hand movement. We expected that the earlier the robot gazed at participants, the shorter the reaction time would be provided that the gaze shift was perceived as a cue to take turns. However, we do not expect this behaviour based on the GOADI theory, because the GOADI theory does not incorporate social cues. In addition, we speculated that the reaction time of participants to imitate would be shorter if a goal object is present than if it is absent.

2 Method

2.1 Task

Participants were requested to play an imitation game with a humanoid Nao robot (Aldebaran Robotics, Fr.) [39]. They

were instructed to imitate the movement of the robot after the robot completed its movement. The robot moved one of its hands to a goal location with or without the goal objects. The robot looked at its hand and back to the participant with four different timings. In total, there were eight conditions. Within each condition, there were four different types of movements to imitate: the robot is pointing either to its ipsi-lateral side or contra-lateral side, and either with its left or right hand. Subsequently, participant's imitation performance was measured by the reaction time (RT) to initiate their movement and movement trajectories using the trakSTAR movement tracking system (Ascension Technology Corporation, USA [40]).

2.2 Participants

Participants were recruited using the participants' database of the Human-Technology Interaction Department of the Eindhoven University of Technology, the social network Facebook and flyers distributed inside the university. In total, 70 participants took part in the experiment, of which 48 were male and 22 were female (mean age 25.34, $SD = 5.61$, range 18–46). Among them, 68 participants were right-handed and 47 already had prior experience with the Nao robot. Participants received € 7.5 as compensation or € 9.5 if they were not affiliated with Eindhoven University of Technology.

2.3 Experiment Setup

The experiment was carried out in the UseLab of the Eindhoven University of Technology. The software Choregraphe [41] was used to generate Nao's body movements. While for voice instructions and other complex behaviours, python scripts were implemented using the Naoqi software development kit [42].

Figure 3 shows a top-down view of the experiment setups. Participant sat face-to-face with the Nao robot. The experimenter observed the participant's performance using a see-through mirror placed behind the participant. The red dashed line indicates the left-right symmetry axis (not present in the real setup). Two grey dots represent the positions of two cups and the grey crosses represent the starting positions of the participant's hands. The distance from the center of the cups to the participants hands and robot hands was equal, which is about 16 cm. The two cups were 31 cm apart with the imaginary red line in the middle. This was the same for the two grey crosses. The radius of each grey cross was 3 cm. The center of the grey cross (the black dot) indicated where participants had to place their index fingertips.

The measurement device, the trakSTAR is a high-accuracy electromagnetic tracking device for short-range motion tracking applications [40]. It composes of a trakSTAR transmitter, two small sensors and electronics unit. The transmitter

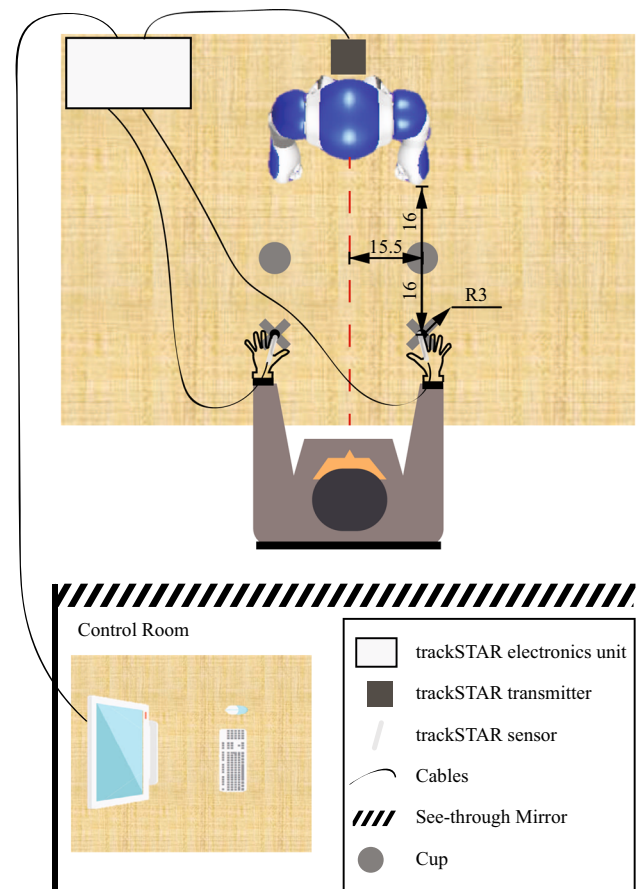


Fig. 3 A top down view of experiment setups (all the measures were in units of cm)

has a maximum range of 90 cm. The sensors have an update rate of 100HZ, and they were attached to participant's index fingertips with skin tapes, in this way the movements can be tracked in real-time. The sensors were calibrated by instructing the participant to put their index fingertips on the starting position and setting the values for X, Y, Z coordinate for both hands to zero. The software is written in C# and custom built.

2.4 Procedure

After the participants received the informed consent and established that they had no further questions, they were asked to sit face-to-face with the Nao robot. The robot was aligned with the participant so that it appeared to look directly at the participant when looking straight ahead. The experimenter then adhered skin tapes to participants' hands, and attached one trakSTAR sensor to each participants' index fingertips. After that, the participants were asked to put their index fingertips onto the starting positions (the center of the two crosses in Fig. 3). After the experimenter checked whether all items were set up correctly (i.e. the goal objects had been put on or removed from the table, the skin tapes

were in good adhesion, etc.), the experiment started from the control room.

The experiment started by a greeting from Nao, and then Nao introduced himself and provided a brief explanation of the experiment: “You are asked to imitate me after I complete my movement, when you are ready to imitate me, please put your index fingertips onto the center of the cross signs on the desk. Every time you finished the imitation, please put your index fingertips back onto the cross signs again.” After the explanation, the experimenter first calibrated the zero-positions of the two trakSTAR channels (corresponding to the participant’s initial positions of the left and right index fingertips). Then a five-trial training session was presented in order to let participants familiarize themselves with the task.

The experiment consisted of two parts, which was counterbalanced. In the first part, the forty-eight trials were done with cups (goal-present) or without cups (goal-absent). After finishing the first part, the participants received a 4-min break. In the meantime, the experimenter would remove or put cups onto the desk. Then, the rest of the experiment followed. Halfway each part, Nao encouraged the participant with a sentence “You are doing great! Keep moving” or “Good job! Keep moving”.

After finishing the last trial, participants were asked to fill in their basic information, which included: age, gender, right-handedness/left-handedness, and prior experience with the Nao robot. The whole experiment lasted about 35 min, and ended with a debriefing with the participants.

2.5 Experiment Design

The experiment was conducted as a two-factor within-subject design, which the presence of the goal object and the timing of the robot’s turn-take cue were manipulated.

2.5.1 Goal Object Manipulation

The goal object was included as one of the within-subject factors and composed of two conditions: Goal object present condition (goal-present) and goal object absent condition (goal-absent). In the goal-present condition, two identical cups were placed on the table, so that the objects only differed by location. Robot would point to one of the two cups on the table. In the goal-absent condition, no cups were placed on the table. The robot pointed to the same spatial locations as when the cups were presented. The order was counterbalanced.

2.5.2 Gaze Timing Manipulation

The timing of the robot’s turn-take cue (i.e. gaze timing) was included as a second within-subject factor. The four gaze conditions are depicted in Fig. 4 (see four grey bars only). In

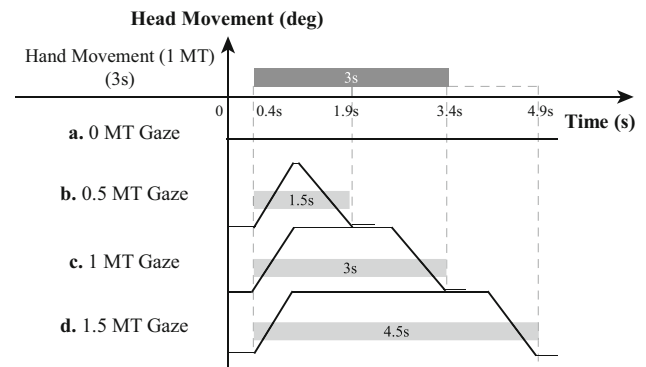


Fig. 4 Four gaze timing conditions with respect to the robots hand movement time (MT). The gaze timings were either be 0 MT (a), 0.5 MT (b), 1 MT (c) or 1.5 MT (d). The grey bars indicated the movement time in seconds. The movement speed is schematically indicated by the solid line

Fig. 4, the robot hand movement time (MT) fixed at 3 s, which included the time needed for moving back and forth. During each trial, the robot first looked at the participant before moving (as shown in Fig. 1). When pointing, it looked at its hand pointing direction (as shown in Fig. 2), after some time, depending on the condition, it looked back at the participant. Four different gaze timings at which the robot looked back at the participant with respect to its hand MT were implemented: (a) 0 MT: the robot always gazed at the participant, and it never gazed to its hand movement direction. (b) 0.5 MT: at the start of the movement, the robot looked at the participant. After half of the hand MT, which was after 1.5 s, the robot looked back at the participant. This meant that when the robot’s hand was still under way, it already looked back at the participant. (c) 1 MT: After completing one hand movement, which was after 3 s, the robot looked back at the participant. (d) 1.5 MT: After one and a half MT, which was after 4.5 s, the robot looked back at the participant.

2.5.3 Gaze: The Head Movement

As the Nao does not have moveable eyes, gaze shifts of the robot were implemented as head movements. From previous studies we know that the perception of eye contact with Nao as a looker, is very similar compared to a human looker [43]. Also, several studies have shown that looking back at the participant is interpreted as a turn-taking cue for various agents [44–46] including the Nao robot [34,47,48]. In Fig. 4 the solid lines indicate the movement on- and off-set. Both hand movement and head movement started at 0.4s due to a “buffering time” for the Nao to start its movement stably and smoothly. The speed with which the head moved was always the same, only the stationary fixation time differed.

2.5.4 Hand Movements

The robot's hand movements consisted of two pointing movements of either hand (left or right hand) by two sides (Ipsi-lateral or contra-lateral side), so four different types as described in Fig. 1: (a) left hand ipsi-lateral (b) left hand contra-lateral (c) right hand ipsi-lateral (d) right hand contra-lateral movement. Note that the left and right are defined with respect to the participants' view. Figure 2 depicts the combination of Nao's hand and gaze movement. As the gaze direction was always corresponding to its hand pointing direction, there were only four possible combinations of head and hand movement.

2.5.5 Dependent Measures

Participants' reaction time (RT) to initiate the imitation was introduced as a dependent variable. Thus, two goals (goal-present or goal-absent) by four gaze timings (0 MT or 0.5 MT or 1 MT or 1.5 MT gaze) conditions were included. For each condition, four different types of robot movements were implemented. All the measurements were repeated three times and pseudo-randomized (i.e. the order was manually modified when the same gestures would appear repeatedly two or three times). In total, there were $2(\text{Goals}) \times 4(\text{Gazes}) \times 4(\text{Movements}) \times 3(\text{repetitions})$ trials, which resulting in 96 trials per participant.

2.6 Data Analysis

The trakSTAR was used to record the movements of the participants, from which the participants' RT, chosen hand, and chosen target were extracted.

2.6.1 Data Preparation

Prior to the statistical analysis, all erroneous trials where the trakSTAR triggered the robot twice were deleted from the data set (8 out of 6720 trials (96 trials \times 70 participants), 0.12%). All erroneous trials where the participants' reaction time exceeded 4s were removed from the data set (12 out of 6712 trials; 0.18%). The data of one participant had to be discarded, because in 23 out of 96 trials (24%) the participant used the correct hand towards the wrong goal. Consequently, 6604 trials were included in the final analysis.

The data analysis was performed using IBM SPSS Statistics 22 (IBM Corporation, 2013). For all statistical analyses the significance level was set to a value of $p = 0.05$ with a confidence interval of 95%.

2.6.2 Reaction Time Extraction Using trakSTAR

After the participant's hands were held for 3 s within a range of 3 cm radius from the starting position, the trakSTAR trig-

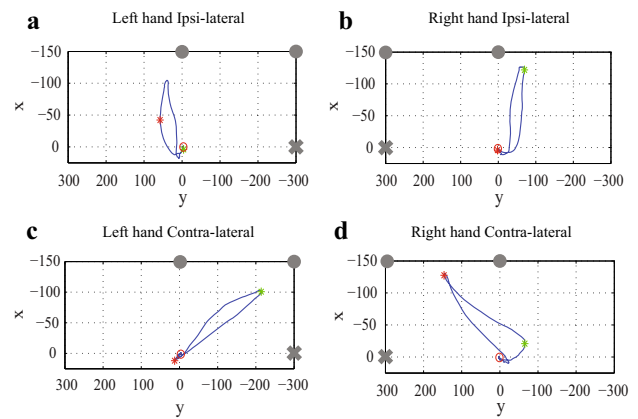


Fig. 5 Top view of hand trajectories of one participant (blue line). The red circle represents the starting position of the imitating hand. The grey cross indicates the starting position of the other hand. The grey dots represent the two target locations. The green and red star indicate the maximum left and right lateral displacements (all units are in mm). The grey dots represents the approximate left and right target position

gered the robot (T1). The robot hand movement lasted 3 s. After the completion of the robot's gesture, the participant started to imitate (T2). When the participants finished imitating, they put their hands back onto the starting positions. This process is repeated for each trial till the end of the experiment.

Which hand the participant chose to imitate the robot was labeled by comparing the maximum displacement of the left and the right hand, and the larger displacement indicated which hand was used.

After determining the used hand, T2 was determined for extracting RT. First, the maximum hand displacement was determined to restrict the search range, as T2 should occur beforehand. Then the stationary position and noise was determined by taking the mean and standard deviation of the first 1s of the movement data, where the hand is still stationary. T2 is defined as the time that the displacement stays above a threshold of the mean + $2 \times$ standard deviation. To determine this numerically, we searched back from the maximum until the first time the distance was below threshold. After finding T2, the RT can be calculated as T2-T1.

2.6.3 The Movement Trajectory Analysis

In Fig. 5, a top view of four typical movement trajectories for one of the participants is shown: (a) left hand ipsi-lateral movement (b) right hand ipsi-lateral movement (c) left hand contra-lateral movement (d) right hand contra-lateral movement.

Which target a participant chose is determined off-line from the maximum lateral displacement: in Fig. 5, the red star indicated the left maximum lateral displacement (LD), and the green star indicated the right maximum lateral displacement (RD). In general, if the maximum lateral dis-

placement exceeds 100 mm we considered the movement to be contra-lateral. For example, in Fig. 5b the maximum lateral displacement of the right hand (red cross) was less than 100 mm, so the movement is ipsi-lateral and the right target was chosen. In Fig. 5d the maximum leftward movement exceeds 100 mm, so the movement is contra-lateral and the left target is chosen. A similar argument applies to a left hand movement (as shown in Fig. 5a, c). However, due to variation between movement trajectories, it was necessary to combine the automatic detection of movement type and target location with visual inspection. For some participants a criterion of 90 mm was more appropriate whereas for other participants the trajectories were extremely curved, so that our simple rule was no longer appropriate. In those cases we manually determined the movement type and target location.

3 Results

3.1 Mirror Symmetrical and Anti-mirror Symmetrical Imitation

First, we determined which hand the participants used to imitate the robot’s movements. Participants could either imitate with the hand on the same side, i.e. as if they are facing a mirror (mirror symmetric imitation), or with their opposite hand (anti-mirror symmetric). More formally, the latter could also be labeled symmetric imitation, but since mirror symmetric imitation is the default behaviour, we used the label anti-mirror symmetric to stress this fact. We found that among 70 participants, 55 (78.6%) adopted ‘mirror symmetric’ imitation, whereas 15 (21.4%) adopted ‘anti-mirror symmetric’ imitation. In Fig. 6 the distribution of participants is shown for correct mirror-symmetric imitation (same hand and same target). As can be seen, the distribution is bi-modal with peaks at close to 0 and 100% without any participants in the

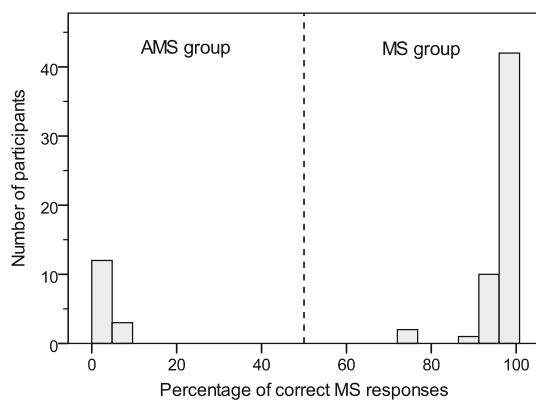


Fig. 6 Histogram of the percentage of correct MS imitations. The ordinate shows the number of participants. Percentages of the MS and AMS groups are much larger and much smaller than 50%, respectively

intermediate range (20–65%). Whether participants use mirror symmetric or anti-mirror symmetric imitation determines which target should be considered as correct. Therefore, we divided our participants into two groups, the ‘mirror symmetric’ (MS) group and the ‘anti-mirror symmetric’ (AMS) group according to their hand use.

It is possible that in doing so a difference between the AMS and MS is introduced. To check this, we conducted a MANOVA with age, gender, right handedness, and experience with Nao as dependent variables and MS/AMS group as independent variable. We found no effect MS/AMS group on any of the dependent variables ($F(4, 64) = 0.311, p = 0.87$).

The correct MS imitation occurs when both the chosen hand and target are chosen mirror-symmetrically. For the AMS group correct imitation occurs when hand and target are chosen anti-mirror symmetrically. This is shown in Fig. 7a, b. As a result both groups have a high ‘correct’ imitation rate (>90%, see below).

3.2 The MS Group

3.2.1 The Error Types

Apart from correct imitations (Fig. 7a), either the hand, the target or both could be incorrect. These ‘errors’ (Fig. 7c) can be subdivided as the ‘Goal Error’ (the chosen hand was correct but the chosen target was wrong), the ‘Means Error’ (the chosen target was correct but using the wrong hand), and

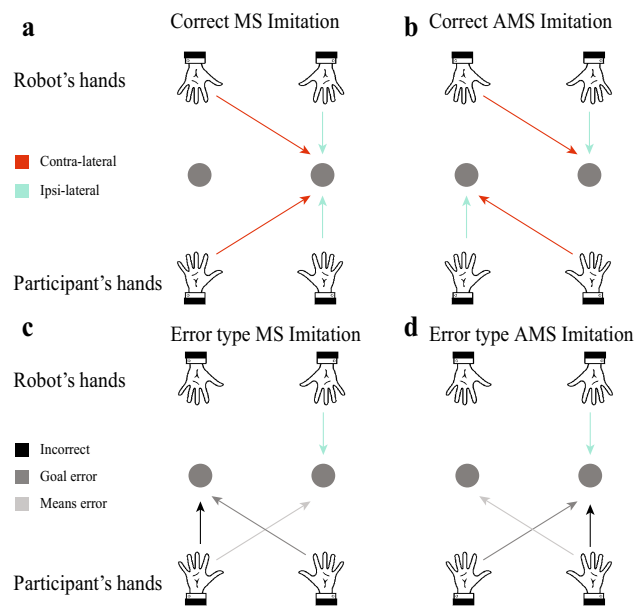


Fig. 7 **a** The correct MS imitation. **b** The correct AMS imitation. A red arrow represents a contra-lateral movement, and a green arrow represents an ipsi-lateral movement. **c** Examples of the error types of the MS imitation. **d** Examples of the error types of the AMS imitation



Fig. 8 Error types of the MS and the AMS group

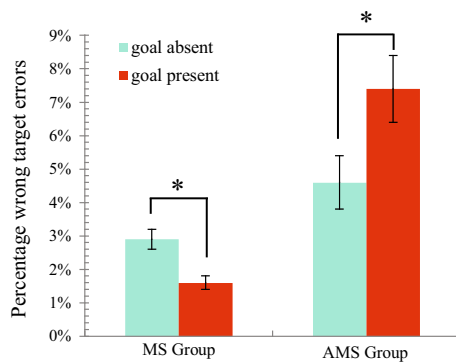


Fig. 9 The effect of goal presence on the percentages of wrong target choices

the ‘Incorrect’ error (both the chosen hand and the chosen target were incorrect).

Figure 8 shows the percentages of correct imitation and the three error types. Most movements are imitated in a mirror-symmetric fashion (97.3%). To compare the different error types, we did a non-parametric test with equal probabilities. We found a significant deviation from the assumption that all error types are equal ($\chi^2(2) = 27.817$, $p < 0.001$). We also compared the goal errors and means errors directly by comparing the observed frequencies of goal and means errors: we found a significantly larger probability of making a goal error than a means error ($\chi^2(1) = 14.098$, $p < 0.001$).

3.2.2 The Effect of Goal Presence on Imitation

When the target object is present, the goal is more salient. In that case, it is expected that the participants chose the wrong target less frequently in the goal-present condition than in the goal-absent condition (both a ‘Goal Error’ and ‘Incorrect’ imitation).

Figure 9 (see the MS group) shows the percentages of the wrong target choice in both goal conditions: when a goal is absent, there were 2.9% wrong target choice, and when a

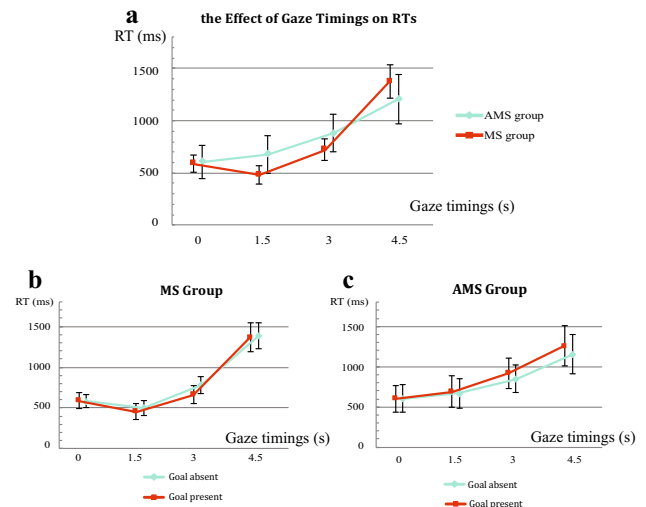


Fig. 10 The effect of gaze timings on RTs, error bars represented standard error mean

goal is present, there were only 1.6% wrong target choice. An independent-samples t-test was run to determine if there were differences in the amount of wrong target choices between the goal-absent and the goal-present conditions in the MS group. We found that the percentage of wrong target choices in the MS group is higher in the goal-absent than in the goal-present condition ($\Delta M = 1.3\% \pm 0.4\%$, $t(4789.906) = 3.114$, $p = 0.002$). We also tested whether the presence of the goal object modified the difference between the percentage of goal errors and means errors. The rationale is that the target object improves goal saliency, so that relatively fewer goal errors are made. We did a Pearson Chi square test of independence between goal error/means error and goal present/absent. We found a highly significant interaction effect ($\chi^2(1) = 10.607$, $p = 0.0015$, $N = 82$) showing that the probability of a goal error over a means error was lower when the goal was present than when it was absent.

3.2.3 The Effect of Goal Presence and Gaze Timing on RT

All the RTs were measured relative to the start of the robot’s hand movement instead of the end of its hand movement. So 3 s (the duration of the robot’s hand movement) were subtracted from the raw RTs. Reaction times shorter than the robots movement onset were removed.

A two-way repeated measures ANOVA was conducted to determine whether there was a statistically interaction effect in RT over different gaze timings and goal manipulation. The data was normally distributed at each interaction effect level, as assessed by a boxplot and the Shapiro–Wilk test ($p > .05$).

First, the main effect of gaze timings is presented. The result is shown in Fig. 10a (red curves). Mauchly’s test of sphericity indicated that the assumption of sphericity had

Table 1 The mean difference between consecutive gaze timings for the MS group

MS group			
	Mean difference	SE	<i>p</i> value
Gaze 0–Gaze 1.5	0.109	0.025	<0.001
Gaze 1.5–Gaze 3	−0.244	0.034	<0.001
Gaze 3–Gaze 4.5	−0.652	0.074	<0.001

been violated, $\chi^2(5) = 122.769$, $p < .0005$. Therefore, a Greenhouse-Geisser correction was applied. The gaze timing elicited statistically significant changes in RT:

$F(1.304, 66.529) = 73.814$, $p < .0005$, with RT decreasing from gaze timing 4.5s ($M = 1.37$, $SD = 1.15$) to gaze timing 3 s ($M = 0.72$, $SD = 0.73$) to gaze timing 1.5s ($M = 0.47$, $SD = 0.63$). However, for gaze timing 0 s ($M = 0.59$, $SD = 0.59$), this decreasing pattern did not occur. Post hoc analysis with a Bonferroni adjustment revealed that RT differed significantly between consecutive gaze timings, although the difference between gaze timings of 0 and 1.5 s differed in the opposite direction (see Table 1).

Second, the main effect of the goal is analyzed. We did not find any statistical difference between the goal conditions. RT decreased from the goal-absent ($M = 0.81$, $SD = 0.74$) to the goal-present ($M = 0.77$, $SD = 0.65$), with $F(1, 51) = 0.5$, $p = 0.483$.

Last, the interaction effect was examined (see Fig. 10b). However, there was no statistically significant interaction in RT between the goal manipulation and gaze timings ($F(2.152, 109.774) = 1.697$, $p = 0.186$.)

3.3 The AMS Group

3.3.1 The Error Types

Figure 7d showed the error types of AMS group. As shown in Fig. 8, in the AMS group, the percentages of correct imitations, goal errors, means errors and the incorrect errors were, 93.58, 3.56, 0.42, 2.44%, respectively. Like the MS group, a significant deviation from the assumption that all error types were equal was also found ($\chi^2(2) = 33.935$, $p < 0.001$). In fact, the participants made significantly more goal errors (3.56%) than means errors (0.42%) ($\chi^2(1) = 35.526$, $p < 0.001$). Note that for the AMS group, a goal error meant that the participants chose the same target as the robot did, which could reflect that the real position of the target was dominant for the AMS group.

3.3.2 The Effect of Goal on Imitation

For the AMS group, the ‘correct’ imitation is to choose the different target as the robot did (Fig. 7b). When the target

is present, we inferred that the participants were prone to choose the same target as the robot. In other words, for the AMS group, we assume that the participants chose the wrong target more often in the goal-present condition than that in the goal-absent condition.

Figure 9 (see the AMS group) shows the percentages of wrong target choices in the goal absent condition (4.6%) and goal present condition (7.4%). The result of independent-samples t-test indicated that the wrong target choice in the AMS group was higher in the goal-present condition ($\Delta M = -2.8\% \pm 1.3\%$, $t(1362.173) = -2.252$, $p = 0.024$). To test whether the presence of the goal object modified the difference between the percentage of goal errors and means errors, we did a Pearson Chi square test of independence between goal error/means error and goal present/absent. We found no significant effect ($\chi^2(1) = 4.215$, $p = 0.072$, $N = 57$).

3.3.3 The Effect of Goal and Gaze Timing on RT

The same RT analysis as the MS group was applied to the AMS group. In the AMS group, Mauchly’s test of sphericity indicated that the assumption of sphericity was not violated, $\chi^2(5) = 10.235$, $p = .07$. The gaze timing elicited statistically significant changes in RT, $F(3, 36) = 14.119$, $p < .001$. Specifically, as shown in Fig. 10a (blue curves): RT decreased from gaze timing 4.5s ($M = 1.21$, $SD = 0.84$) to gaze timing 3 s ($M = 0.89$, $SD = 0.63$) to gaze timing 0.5s ($M = 0.68$, $SD = 0.64$) to gaze timing 0 s ($M = 0.60$, $SD = 0.59$). Post hoc analysis with a Bonferroni adjustment revealed that RT was not significantly different between consecutive gaze timings. We only found a significant difference between a gaze timing of 0 and 3 s (mean difference = -0.277 , 95% CI [$-0.483, -0.075$], $p = 0.006$), and a gaze timing 1.5s and 4.5s (mean difference = -0.528 , 95% CI [$-0.962, -0.095$], $p = 0.014$).

We did not find a significant main effect of goal condition for the AMS group (goal-absent: $M = 0.82$, $SD = 0.65$; goal-present: $M = 0.87$, $SD = 0.68$, with $F(1, 12) = 0.274$, $p = 0.610$).

Last, the interaction effect was examined (see Fig. 10c). However, there was no statistically significant interaction in RT between the goal manipulation and gaze timings ($F(3, 36) = 0.487$, $p = 0.693$).

4 Discussion and Conclusions

In the current study, we used imitation to investigate whether participants treat the humanoid Nao robot as an ‘intentional agent’. Specifically, we verified whether the GOADI theory, which was initially applied to HHI, is also applicable to HRI. This provides us with insights about whether humans imitate

robot movements in the same way as human movements. The GOADI theory does not make any predictions about whether people treat robots in a social way. Therefore, we also used an eye gaze movement that mimicked a turn-taking cue. It is known that this is a very effective social cue for artificial agents [34]. Since participants were instructed to imitate the robot's hand movement as soon as it was completed, the gaze cue is irrelevant for the imitation task. Only when the gaze shift is interpreted as a social cue, we would expect an effect. By combining these measures in a single task, we were able to assess both imitation behaviour and sensitivity to a social cue.

The results showed that the GOADI theory was partly supported for HRI. In general, we found there were two types of imitation patterns, mirror-symmetric (MS) imitation and anti-mirror-symmetric (AMS) imitation. Most participants (78.6%) preferred MS imitation. This is in accordance with the findings that humans tend to imitate in a MS fashion rather than in an AMS fashion [16, 24, 49]. Surprisingly, other studies did not report any data on anti-mirror symmetric imitation. It could be that the AMS group was discarded as erroneous data, or it could be that percentage of participants using anti-mirror symmetric imitation is larger when imitating a robot than a human. The latter would point to the interpretation that robotic movements are inherently more ambiguous, so that participants are less confident about what to imitate.

In the MS group, the presence of the goal reduced the number of goal errors, a finding which is consistent with the GOADI theory. However, participants made more goal errors than means errors, which means that participants tended to imitate the hand movement of the robot instead of the goals. This is not in support of the GOADI theory. For the AMS group, participants not only made more goal errors than means errors, but the number of goal errors also increased when the goal object was present. Here, however, the interpretation is different from the MS group. For the AMS group, a correct imitation is to choose the other goal than that of the robot (see Fig. 7b). Thus, the more goal errors are made, the more participants move toward the goal of the robot. Thus it could be that these errors are made because there is a tendency to share the goal object with the robot. This would be in agreement with the GOADI theory; however, there is no statistical support for this argument because it is only for 21.4% of the participants. The same conclusion is obtained from visual inspection of the trajectory data (see Fig. 5). Most participants imitated the movement path as presented by the robot (i.e. when imitating a contra-lateral movement, they didn't go directly as a straight line towards the cup, but followed a C-shaped trajectory as the robot did). This is somewhat surprising since human movements could not be identical to those of the robot. The robot was crouched on the table with the arms down, whereas participants started with their hands on the table in front of them. If anything,

we would have expected participants to shift their decision of what to imitate to a higher level of abstraction, that is, imitating the goal. This is not what we found, however. Participants tried to imitate the movement trajectories as best as they could given the constraints from the environment (the table). Therefore, we reason that most of participants used a lower level representation than the high level goal representation to imitate.

For the effect of the presence of a goal object on the RT to initiate imitation, we expected that the participants would react faster when the goal is present. However, for both the MS and AMS group, the presence of the goal object didn't help the participants to react faster. Again this suggests that participants paid more attention to imitating the hand movement of the robot than the goal.

The turn-taking indicated by the eye gaze cue was quite effective. Generally, the participants' reaction time to start imitating increased with gaze timing, even though this cue was irrelevant for the imitation task. The earlier the robot looked at the participants with respect to its movement, the earlier the participants started imitating. The 'gaze timing 0 s' deviated from this pattern. A possible explanation is that the robot never looked away in this condition. Thus, the gaze time is not well-defined, although we had arbitrarily set it to 0 s. If we disregard the no gaze time condition, the effect of gaze timing follows a more or less linear trend with a slope of about 0.3 (0.9 s shift in RT for 3 s difference in gaze timing) for the MS group and about 0.16 for the AMS group (see Fig. 10a). The effect of the turn-taking cue did not depend on the manipulation of the goal object, because there was no interaction effect between these variables. The range of gaze timings was huge (3 ± 1.5 s). Presumably, the effectiveness of the gaze cue will deteriorate when it is delayed too much, or when the movement is too fast or slow. However, good quantitative data on turn-take timing is still missing.

According to Wykowska et al. [6] an *intentional stance* is required before people can attribute mental states to agents, and it was shown that beliefs about agency affect the effectiveness of a gaze cue [4, 6]. The effects of gaze timing on the onset of imitation in our study are much larger than those observed using a gaze cueing paradigm, which is about 15 ms [4]. This suggests that participants adopted the intentional stance in our experiment and regarded the Nao robot as a social agent. From our data it is not possible to distinguish whether participants really believed the robot to be a social agent or that they believed it to be operated by a social agent. To be able to distinguish between the two possibilities we would have to manipulate the robot's agency directly. Another explanation for the large effect of gaze timing is that the 'gaze cue' acted as a go-signal. In this case an auditory beep would have produced similar results. It is hard to say to what extent the go-signal effect plays a role in our data. As a counterargument it may be noted that a robotic gaze

shift does not automatically imply a turn-taking cue [5]. If so, the gaze shift would be more similar to a posture shift and would not act as a go-signal. Furthermore, a go-signal does not really explain why the effect of gaze timing has a gain of at most 0.3. If an auditory beep would have been used as a go-signal one would expect a gain much closer to 1. Still a direct comparison between an auditory cue, a gaze shift, and a turn-take cue would be needed to sort out whether a turn-take cue is merely a go-signal.

The GOADI theory holds that people imitate observed actions based on a hierarchical organisation of (sub)goals. As a result, when the goal object is present, the goal saliency is increased and the goal receives higher priority when imitating. We did indeed observe fewer goal errors when the goal object was present than when it was not. So, it seems HRI is similar to HHI in this respect. We also found that overall fewer means errors were made than goal errors, which suggests that the participants imitated the means before the goal. One possible explanation is that the saliency of the action goals was still rather low: the robot did not touch the objects and the goal objects were identical. The ambiguity of the goals may have diminished goal saliency, but it is unclear whether this applies to the present study. Our results show very low error rates, so there is no reason to believe that the target that was being pointed at was unclear. Earlier studies also used very similar target objects (e.g. identical dots in [16], ears in [24], identical mugs with different prints in [26]), so it seems unlikely that the identical appearance of the target objects could have had a large effect. However, further research would be needed to clarify this issue. For HHI, it was found that people shift to imitating the movement trajectories when the saliency of the goals is low [50–52]. Indeed we found that many subjects reproduced the curvedness of the robot's pointing movements even though this was only necessary to prevent the robot from hitting its knee. Because the object was not touched, this could explain why there were fewer means errors than goal errors. On the other hand, Gazzola et al. [8] found similar mirror neuron system (MNS) activation for human and robotic arm movements even though the latter used a constant velocity and only one degree of freedom. So, future studies should examine the effect of deviating kinematics and movement trajectories on the saliency of action goals. Another complication stems from [8], where it was shown that highly repetitive robotic movements do not trigger the MNS. In our study participants experienced the four different arm movements 32 times, but they were accompanied with four different gaze behaviours and two goal presence conditions (each condition was repeated three times). It is possible that the repetitiveness could have had a negative effect on the responsiveness of the MNS, thereby reducing the saliency of the associated action goals. On the other hand, the repetitiveness in our study is quite similar to the HHI studies (e.g. [16,26]). The experiment

was also fast-paced (95 trials in 35 min.), which makes the task suitable for assessing relatively low-level mechanisms like the MNS, which is presumed to play an important role. In any case it would be interesting to investigate the GOADI theory further when the task is less repetitive and more complex. It is also possible that participants were unsure about the relative importance of what to imitate, because it is a robot. This could then lead to the selective imitation of the movement trajectory as suggested by Lyons [25]. Viewed in this way, we can speculate that the movements of the robot are less predictable, because our internal simulation of observed intentional actions does not work well for robots. The resulting uncertainty would then make the inference of action goals rather difficult [21]. So simulation theory is able to explain why action goals of robots are not always imitated. However, simulation theory and GOADI do not make any claims about when to imitate observed movements. In this study we have shown that when participants imitate strongly depends on the timing of a turn-taking cue. The integration of social cues on the decision of when to imitate what, is largely unexplored, however.

To summarize, we found partial support for the GOADI theory: the presence of a goal object reduced the number of goal errors that participants make (in the MS group), but the number of goal errors is larger than the number of means errors. This may be due to the fact that the robot's movements still deviate from real human movements, which action goal inference more difficult. We also found that gaze timing is an effective turn-taking cue, which suggests that participants treat the robot as a 'social agent'.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Breazeal C (2004) Social interactions in HRI: the robot view. *IEEE Trans Syst Man Cybern C Appl Rev*. doi:[10.1109/TSMCC.2004.826268](https://doi.org/10.1109/TSMCC.2004.826268)
2. Fiore SM, Wiltshire TJ, Lobato EJC, Jentsch FG, Huang WH, Axelrod B (2013) Toward understanding social cues and signals in human robot interaction: effects of robot gaze and proxemic behaviour. *Front Psychol Cognit Sci* 4(859):1c15
3. Scassellati B (2002) Theory of mind for a humanoid robot. *Auton Robots* 12(1):13–24
4. Wiese E, Wykowska A, Zwicker J, Müller H (2012) I see what you mean: How attentional selection is shaped by ascribing intentions to others. *PLoS ONE* 7(9):e45391. doi:[10.1371/journal.pone.0045391](https://doi.org/10.1371/journal.pone.0045391)
5. Wykowska A, Kajopoulos J, Obando-Leitón M, Chauhan SS, Cabibihan JJ, Cheng G (2015) Humans are well tuned to detecting

- agents among non-agents: examining the sensitivity of human perception to behavioural characteristics of intentional systems. *Int J Soc Robot* 7(5):767–781
6. Wykowska A, Wiese E, Prosser A, Müller HJ (2014) Beliefs about the minds of others influence how we process sensory information. *PLoS ONE* 9(4):e94339. doi:[10.1371/journal.pone.0094339](https://doi.org/10.1371/journal.pone.0094339)
 7. Alaerts K, Swinnen SP, Wenderoth N (2010) Observing how others lift light or heavy objects: Which visual cues mediate the encoding of muscular force in the primary motor cortex? *Neuropsychologia* 48(7):2082–2090
 8. Gazzola V, Rizzolatti G, Wicker B, Keysers C (2007) The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* 35(4):1674–1684
 9. Hegel F, Krach S, Kircher T, Wrede B, Sagerer G (2008) Theory of mind (ToM) on robots: a functional neuroimaging study. In: Proceedings of the 3rd ACM/IEEE international conference on human robot interaction (HRI), pp 335–342
 10. Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behav Brain Sci* 1(04):515–526
 11. Gallese V, Goldman A (1998) Mirror neurons and the simulation theory of mindreading. *Trends Cognit Sci* 2(12):493–501
 12. Gazzola V, Aziz-Zadeh L, Keysers C (2006) Empathy and the somatotopic auditory mirror system in human. *Curr Biol* 16:1824–1829
 13. Rizzolatti G, Craighero L (2004) The mirror-neuron system. *Annu Rev Neurosci* 27:169–192
 14. Umiltà MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, Rizzolatti G (2001) I know what you are doing: a neurophysiological study. *Neuron* 31:155–165
 15. Brass M, Bekkering H, Prinz W (2001) Movement observation affects movement execution in a simple response task. *Acta Psychol* 106(1–2):3–22
 16. Wohlschläger A, Bekkering H (2002) Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Exp Brain Res* 143(3):335–341
 17. Ribeiro AC (2003) Do mirror neurons support a simulation theory of mind-reading? Graduate philosophy conference, San Diego, p 12
 18. Sebanz N, Bekkering H, Knoblich G (2006) Joint action: bodies and minds moving together. *Trends Cogn Sci* 10(2):70–76
 19. Williams JHG, Whiten A, Suddendorf T, Perrett DI (2001) Imitation, mirror neurons and autism. *Neurosci Biobehav Rev* 25:287–295
 20. Billard AG, Calinon S, Guenter F (2006) Discriminative and adaptive imitation in uni-manual and bi-manual tasks. *Robot Auton Syst* 54(2006):370–384
 21. Cuijpers RH, van Schie HT, Koppen M, Erlhagen W, Bekkering H (2006) Goals and means in action observation: a computational approach. *Neural Netw* 19(3):311–322
 22. Erlhagen W, Mukovskiy A, Bicho E, Panin G, Kiss C, Knoll A, van Schie H, Bekkering H (2006) Goal-directed imitation for robots: a bio-inspired approach to action understanding and skill learning. *Robot Auton Syst* 54(5):353–360. doi:[10.1016/j.robot.2006.01.004](https://doi.org/10.1016/j.robot.2006.01.004)
 23. Oberman LM, McCleery JP, Ramachandran VS, Pineda JA (2007) EEG evidence for mirror neuron activity during the observation of human and robot actions: toward an analysis of the human qualities of interactive robots. *Neurocomputing* 70:13–15. doi:[10.1016/j.neucom.2006.02.024](https://doi.org/10.1016/j.neucom.2006.02.024)
 24. Bekkering H, Wohlschläger A, Gattis M (2000) Imitation of gestures in children is goal-directed. *Q J Exp Psychol* 53A:53–164
 25. Lyons DE (2009) The rational continuum of human imitation. In: Pineda JA (ed) *Mirror neuron systems*. Humana Press, New York, pp 77–103
 26. Wohlschläger A, Grattis M, Bekkering H (2003) Action generation and action perception in imitation: an instance of the ideomotor principle. *Philos Trans R Soc Lond B Biol Sci* 358:501–516
 27. Gleissner B, Meltzoff AN, Bekkering H (2000) Childrens coding of human action: cognitive factors influencing imitation in 3-year-olds. *Dev Sci* 3:405–414
 28. Ferreira JF, Dias J (2014) Attentional mechanisms for socially interactive robots a survey. *IEEE Trans Auton Ment Dev* 6(2):110–125. doi:[10.1109/TAMD.2014.2303072](https://doi.org/10.1109/TAMD.2014.2303072)
 29. Williams M, Abidi SR, Gärdenfors P, Wang X, Kuipers B, Johnston B (2013) Interpreting robot pointing behaviour. In: Proceedings of ICSR, pp 148–159
 30. Belopolsky AV, Olivers CNL, Theeuwes J (2008) To point a finger: attentional and motor consequences of observing pointing movements. *Acta Psychol* 128:56–62
 31. Kleinke CL (1986) Gaze and eye contact: a research review. *Psychol Bull* 100(1):78–100
 32. Khoramshahi M, Shukla A, Raffard S, Bardy BG, Billard A (2016) Role of gaze cues in interpersonal motor coordination: towards higher affiliation in human–robot interaction. *PLoS ONE* 11(6):e0156874. doi:[10.1371/journal.pone.0156874](https://doi.org/10.1371/journal.pone.0156874)
 33. Fischer K, Jensen LC, Kirstein F, Stabinger S, Erkent Ö, Shukla D, Piater P (2015) The effects of social gaze in human–robot collaborative assembly. In: Tapus A, André E, Martin J, Ferland F, Ammi M (eds) *Lecture notes in computer science, social robotics*, 9388th edn. Springer, Berlin, pp 204–213 doi:[10.1007/978-3-319-25554-5_21](https://doi.org/10.1007/978-3-319-25554-5_21)
 34. Bohus D, Saw CW, Horvitz E (2014) Directions robot. In: The-wild experiences and lessons learned, proceedings of the 2014 international conference on autonomous agents and multi-agent systems, Paris
 35. Ito M, Tani J (2004) Joint attention between a humanoid robot and users in imitation game. In: IEEE international conference on development and learning, pp 277–282
 36. Kose-Bagci H, Dautenhahn K, Nehaniv CL (2008) Emergent dynamics of turn-taking interaction in drumming games with a humanoid robot. In 17th IEEE international symposium on robot and human interactive communication 2008 (RO-MAN'08). IEEE, p 346–353
 37. Jokinen K (2010) Non-verbal signals for turn-taking and feedback. In: Proceedings of the seventh international conference on language resources and evaluation (LREC)
 38. Gu E, Badler N (2006) Visual attention and eye gaze during multiparty conversations with distractions. In Proceedings of the international conference on intelligent virtual agents
 39. NAO Documentation—Aldebaran 2.1.4.13 documentation. (n.d.). Retrieved 16 Sept 2016. http://doc.aldebaran.com/2-1/home_ nao.html
 40. Ascension Technology Cooperation (2011) 3D guidance trakSTAR 2TM installation and operation guide
 41. Choregraphe User Guide—Aldebaran 2.1.4.13 documentation. (n.d.). Retrieved 16 Sept 2016. <http://doc.aldebaran.com/2-1/software/choregraphe/index.html>
 42. NAOqi Framework—Aldebaran 2.1.4.13 documentation. (n.d.). Retrieved 16 Sept 2016. <http://doc.aldebaran.com/2-1/ref/index.html>
 43. Cuijpers RH, Pol D van der (2013) Region of eye contact of humanoid Nao robot is similar to that of a human. In Herrmann G, Pearson MJ, Lenz A, Bremner P, Spiers A, Leonards U (eds) *Proceedings of the 5th international conference on social robotics, ICSR 2013, October 27–29, 2013, Bristol, UK, Lecture notes in computer science*, vol. 8239. Springer, Heidelberg, pp 280–289
 44. Chao C, Thomaz AL (2010) Turn-taking for human–robot interaction. In: Proceedings of the AAAI symposium on dialog with robots

45. Kuno Y, Sadazuka K, Kawashima M, Yamazaki K, Yamazaki A, Kuzuoka H (2007) Museum guide robot based on sociological interaction analysis. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI '07). ACM, New York, pp 1191–1194. doi:[10.1145/1240624.1240804](https://doi.org/10.1145/1240624.1240804)
46. Yamazaki A, Yamazaki K, Kuno Y, Burdelski M, Kawashima M, Kuzuoka H (2008) Precision timing in human–robot interaction: coordination of head movement and utterance. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI '08). ACM, New York, pp 131–140. doi:[10.1145/1357054.1357077](https://doi.org/10.1145/1357054.1357077)
47. Bohus D, Horvitz E (2011) Multiparty turn taking in situated dialog: study, lessons, and directions. In: Proceedings of SIGdial2011, Portland
48. van Schendel JA, Cuijpers RH (2015) Turn-yielding cues in robot–human conversation. In: Salem M, Weiss A, Baxter P, Dautenhahn K (eds) Proceedings of the 4th international symposium on new frontiers in human–robot interaction, Canterbury, pp 21–22
49. Perra O, Gattis M (2008) Reducing the mapping between perception and action facilitates imitation. *Br J Dev Psychol* 26:133–144
50. Elsner B, Pfeifer C (2012) Movement or goal: goal salience and verbal cues affect preschoolers' imitation of action components. *J Exp Child Psychol* 112(2012):283–295. doi:[10.1016/j.jecp.2012.02.010](https://doi.org/10.1016/j.jecp.2012.02.010)
51. Gattis M, Bekkering H, Wohlschläger A (2002) Goal-directed imitation. In: Meltzoff AN, Prinz W (eds) *The imitative mind: development, evolution, and brain bases*. Cambridge University Press, Cambridge, pp 183–205
52. Wagner L, Yocom AM, Greene-Havas M (2008) Children's understanding of directed motion events in an imitation choice task. *J Exp Child Psychol* 100:264–275

Yunqing Bao is a PDEng trainee (Professional Doctorate in Engineering) in User System Interaction at Eindhoven University of Technology (TU/e), in the Department of Industrial Design. She received her master degree in Human-Technology Interaction from TU/e. She has a passion and perseverance in building a bridge between innovative technologies and users. Her current research interests are social robotics, wearable technology, usability and user experience.

Raymond H. Cuijpers is associate professor cognitive robotics and human–robot interaction. He graduated in Applied Physics at the Eindhoven University of Technology (NL) in 1996. He received his Ph.D. degree in Physics of Man from Utrecht University in 2000. He did a postdoc on the role of shape perception on human visuo-motor control at Erasmus MC Rotterdam. In 2004 he did a second postdoc at the Radboud University Nijmegen in the context of the European FP6 project called Joint Action Science and Technology (JAST), where he developed computational models of joint action. In 2008 he was appointed as assistant professor at the Eindhoven University of Technology (NL). He has been project coordinator of the European FP7 project Knowledgeable Service Robots for Aging (KSERA) that was successfully completed in 2013. In this context he supervised post-doctoral and Ph.D. students, who studied how robots can autonomously interact with humans for helping older persons. In 2014 he became associate professor Cognitive Robotics and Human–Robot interaction.