Do people understand a humanoid robot's 'mind'?

Bao, Y.

Award date: 2014

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
Do people understand a humanoid robot’s ‘mind’?

by Yunqing Bao

identity number 0826717

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:
Dr. Ir. Raymond Cuijpers
Prof. Dr. Wijnand IJsselsteijn
ACKNOWLEDGEMENTS

After several months work, I’m finally here! Writing this quite relaxed section. Looking back in March, I remembered in one of the first few meetings, I had a sudden nosebleed, and it dropped onto my meeting documents. In retrospect, this might be an early sign that tried to tell me it would be a challenging and demanding work. And it was. But, it was really worthy. I learned a lot that I wasn’t able to learn from the class or out of the class, and I quite enjoyed the whole process. I couldn’t have done all my works without the support of all those people whom I would like to express my great gratitude:

My first ‘super’visor, Raymond, who’s got an exceptionally large mind for solving various puzzles, technically, and non-technically, in a really short time, you taught me a lot, shared your ‘tricks’ a lot, and willing to help me a lot no matter whether there’s a meeting scheduled. So an especially big ‘Hartelijke dank’ to you! Wijnand, who gave me very inspiring ideas on the crossroads of the study, thanks! Martin, and Jan Roelof, who are available for my programming questions, thanks!

My friends in Eindhoven, who accompanied with me in- and outside of the Meta-forum, giving me positive and ‘negative’ encouragements. I enjoyed every ‘LOL moments’ with you guys. And my ex- and current officemates, the ‘office golf tournament’ was really fun. All of you make my life in the Netherlands more colorful.

Finally, to my parents who I left at my home country, but mentally supported me all the time: Although you only knew that I was doing something with a cute robot, sometimes when I encountered difficulties, you both ‘encouraged’ me a lot (you said that after I finished the study, you will cook me a lot of delicious food that I couldn’t have in the Netherlands, and let me get rid of the bread here for a while). Well, this is quite a big encouragement for me, and see! Your encouragement has worked. I think you can start to think about the recipe now.
ABSTRACT

Humanoid robots are expected to exist and 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 smooth and natural way. One important aspect for a successful human-robot interaction (HRI) lies in whether people regard the robot as a potential social agent. Interacting with a potential social robot should facilitate people to have 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. It was unclear whether people can interpret the robot’s actions in the same way because the robot has kinematics that substantially differs from those of humans. Imitation (Human imitating the robot’s actions) provides us an intuitive solution for solving this puzzle, because it is closely about action interpretation from others. In the imitation study of humans, the theory of goal-directed imitation (GOADI) holds that the imitator tends to imitate the action goals instead of action means when the goal object is present. However, when the goal object is absent, people tend to imitate the action means. We want to verify whether the GOADI theory applies to HRI by varying the presence and absence of the goal object. This will provide us motor aspects about how human interpret the robot’s actions. However, even though people can extract the goal from the robot’s action and imitate the robot’s action similar to when imitating a human, it doesn’t mean that people treat the robot as a social agent. Because the GOADI theory cannot predict how people react to social cues, we added a social cue. Specifically, we added the turn-taking cue indicated by eye gaze, which can effectively regulate HHI. We want to verify whether it can also effectively regulate HRI. We manipulated the turn taking cue by changing the gaze timing at which the robot looks at people with respect to its hand movement. We expect that the earlier the robot gazes at people, the reaction time for them to start to imitate will be shorter. Consequently, in combination of the GOADI theory and the social turn-taking cue, we can further claim that people can treat the robot as a potential social agent. The results show that the presence of goal object affected people’s imitation pattern; however, most people tend to match the action means rather than the goals of the robot’s action. On the other hand, the social turn-taking cue regulated by eye-gaze is effective. Potential reasons for these finds and directions for further research are discussed.

Key words: the GOADI theory, imitation, human-robot interaction, turn-taking, eye-gaze.
# Table of Contents

## INTRODUCTION
- Action interpretation in imitation ................................................................. 10
- Action observation in imitation ................................................................. 11
- Theory of goal-directed imitation ............................................................... 12
- Social cues in imitation ................................................................. 13
  - Joint attention mechanism ............................................................... 14
  - Turn-taking ................................................................. 14
- The current study ................................................................. 15

## METHOD
- Task ................................................................. 18
- Participants ................................................................. 18
- Experiment design ................................................................. 18
- Experiment setup ................................................................. 20
- Manipulation of hand movement and gaze timing ................................................................. 22
- Measures ................................................................. 24
  - Data extraction using trakSTAR ................................................................. 24
  - Preliminary analysis ................................................................. 25
  - Movement pathways extraction ................................................................. 25
- Procedure ................................................................. 28
- Data analysis ................................................................. 29

## RESULTS
- HAND USE DATA ................................................................. 31
  - ‘Mirror symmetric’ Group and ‘anti-mirror symmetric’ Group ................................................................. 31
  - The general imitation patterns by two groups ................................................................. 32
  - The effect of goal on hand switch ................................................................. 34
  - The effect of robot movement and goal manipulation on hand switch ................................................................. 35
  - The effect of goal on the target switch ................................................................. 36
  - The effect of goal on movement pathways ................................................................. 37
  - The hierarchy of imitation ................................................................. 42
- RT DATA ................................................................. 44
  - The effect of goal and gaze timing ................................................................. 44

## DISCUSSION ................................................................. 48
INTRODUCTION

Humanoid robots are expected to exist and 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 smooth and natural way. One important aspect for a successful human-robot interaction (HRI) lies in whether people regard the robot as a potential social agent (Breazeal, 2004; Fiore, Wiltshire, Lobato, Jentsch, Huang, & Axelrod, 2013). Interacting with a potential social robot should facilitate people to have a better understanding of the robot’s actions and intentions. Therefore, as emphasized by Scassellati (2002), it’s crucial to focus on how people naturally adopt an ‘intentional stance’ (Dennett, 1987) (as cited in Scassellati, 2002), and interpret the behavior 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 (Alaerts, Swinnen, & Wenderoth, 2010). For instance, when you see somebody stretching out his hand toward a cup, you don’t just see a hand movement trajectory; but, you also infer the goal of reaching that cup, like, drinking. When you see somebody turning his head towards you, you don’t just see a trajectory of head movement; you also infer the intention to get your attention. However, in terms of HRI, though the humanoid robot looks similar to a human, their movements have kinematics that still substantially differs from those of humans (Gazzola, Rizzolatti, Wicker, and Keysers, 2007). Imagining if a humanoid robot is stretching out its hand toward a cup, or turning its head to look at you, will you still understand its intention, or will you just perceive those movements as trajectories moving through space? The answer depends on whether you treat a robot as a social agent or just a machine. It is still unclear whether humans form mental representations of the robot’s actions which allow them to predict the robot’s behavior and then generate a matching motor response similar to that in HHI (Hegel, Krach, Kircher, Wrede, Sagerer, 2008).

Imitation, here referring to a follower (human) matching his/her own movements to a presenter (the robot), provides an intuitive solution for solving this puzzle. The imitation mechanism that underlies HRI is of great interest because it’s closely about action interpretation from others. Moreover, imitation provides an avenue of nonverbal communication through the language of gestures (Meltzoff, 2002), which allows the investigation of social cues in HRI.
Therefore, in the current study, we want to turn to imitation as a substrate for learning what kind of (social) cues can facilitate people to easily interpret the robot’s intention, thus, gaining a successful HRI.

**Action interpretation in imitation**

Before looking into the complex ‘how’ imitation occurs, we first consider ‘what’ is involved in an act of imitation. In order to successfully match one’s movements to other’s action, we first need to visually perceive the action, namely, action observation. After perceiving the action, interpreting and representing of the action comes into play. Last, we execute the motor output to reproduce other’s action (Meltzoff, 2002). We will first look into the essence, how action interpretation comes into play.

There are two main views on how people interpret observed action from others. Theory of mind (ToM) holds that people interpret observed behavior from other people according to a generic model of human behavior (Premack & Woodruff, 1978). ToM is a key capability for gaining successful social interaction in HHI. While simulation theory (ST) claims that people simulate the observed actions in the brain using his/her own action system to make predictions from which goals are inferred (Gallese & Goldman, 1998). Neural evidence partly in support of ST was reported by Gallese, Fadiga, Fogassi & Rizzolatti (1996) and Gallese & Goldman (1998): the discovery of ‘mirror neurons’ in the monkey ventral premotor cortex, which discharge, both when the monkey itself performs specific goal-directed hand movements, and when it observes another individual performing the same movements. According to (Rizzolatti & Craighero, 2004; Gazzola, Aziz-Zadeh, & Keysers, 2006; Gazzola et al, 2007), mirror neurons also exist in human brain. Therefore, when someone carries out a goal-directed motor action, for example, when I grasp a mug, those mirror neurons will fire in my brain. The interesting thing is, those same (class of) mirror neurons will also fire when I see another person doing the same action. Therefore, the mirror neuron was seen as biological evidence of ST, as the mirror neurons appear to simulate the brain state of another (Ribeiro, 2003). However, this neurological evidence supporting ST didn’t necessarily rule out ToM, as ToM is a key capability for achieving successful social interaction, which ST cannot fully explain. Current views include the possibility that both ST and ToM are at work in humans (Williams, Whiten, Suddendorf & Perrett, 2001). It was shown that mirror neurons are involved in action understanding (Umiltà, Kohler, Gallese, Fogassi, Fadiga, Keysers, Rizzolatti, 2001), and imitation as well (Brass, Bekkering, Prinz, 2001; Wohlschlager &
Bekkering, 2002). According to the latter, the behavioral evidence showed that there is a decrease in the latencies of actions by observing matching actions.

**Action observation in imitation**

Not only action interpretation plays an important role in imitation, but also action observation. In previous studies (Gallese *et al*, 1996; Gallese & Goldman, 1998), it was shown that the most effective visual stimuli that trigger mirror neurons were the goal-directed actions, in which the experimenter’s hand interacted with objects. As indicated in (Wohlschläger & Bekkering, 2002), objects act as a goal to drive human imitation behavior. Cuijpers, Schie, Koppen, Erlhagen & Bekkering (2006) stated that goals are considered to be important for action observation since they allow the observer to copy the goal of the action without the need to use the exact same means. And the importance of being able to use different action means becomes evident when the observer and the observed actor have different bodies as in HRI (Cuijpers *et al*, 2006). Gazzola *et al* (2007) also suggested that the goal of the action might be more important for mirror neuron activations than the way in which the action is performed. This is supported by the fact that mirror neuron activity is not much reduced when an observer observes both human hand and 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. In accordance with (Cuijpers *et al*, 2006; Gazzola *et al*, 2007), Bekkering, Wohlschläger & Gattis (2000) provided some behavioral evidence that people prefer to choose the goal object over the correct means in goal-directed movement: for example, the human presenter was sitting face-to-face to the follower, and the presenter’s movement is either contra-laterally (using the right hand to touch the left ear or using the left hand to touch the right ear) or ipsi-laterally (using the right hand to touch the right ear or using the left hand to touch the left ear), the followers were asked to imitate the presenter’s movement as if when they look in a mirror. It turned out that the followers made many ipsi-lateral movements towards the correct object in terms of imitating a contra-lateral movement (e.g. when imitating use left hand to touch the right ear, the follower frequently make ‘errors’ and imitate it with their right hand to touch the right ear). Therefore, the correct movement pathway is ‘sacrificed’ to achieve the correct target in a goal-directed movement. Likewise, Lyons (2009) proposed a revised concept of imitation, which is emulation, i.e. reproducing a model’s goal using means of one’s own devising.
Theory of goal-directed imitation

Some researchers (Bekkering et al., 2000; Wohlschläger, Grattis & Bekkering, 2003) studied this object-oriented imitation pattern by varying 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, the goal objects (ears) are always present. So they adapted ‘hand touch one of two ears’ experiment to the ‘hand touch one of two dots on the table’, since the goal object, the dots can be manipulated as presence and absence. It has been shown that when the dots are present, dot selection becomes the highest goal of imitation. However, when the dots are absent, people tend to use the correct hand and/or movement pathway as the highest goal of imitation. Moreover, they also found systematic error patterns in imitation similar to the touching ear experiment: people touch the dot correctly; nevertheless, they use the ‘incorrect’ hand or movement path to imitate.

On the basis of these systematic error patterns, the theory of goal-directed imitation (GOADI) was proposed (Wohlschläger et al., 2003). 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 pathways (Gleissner, Meltzoff & Bekkering, 2000). When the goal object is present, target object selection becomes the highest goal of imitation. However, when the goal object is absent, people tend to use the correct hand and/or movement pathway 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, it has not been shown whether it also applied to HRI. Verifying the GOADI theory in HRI is quite indicative. If people make systematic imitation errors similar to HHI, means they imitate the goal of the robot’s action rather than the means of the robot’s action. We can claim that people can extract the main goal and simulate the observed robot’s movements in the brain using their own action system, and generate a matching motor response similar to that in imitating humans.

However, even though people can extract the main goal of a robot’s movement, this is not sufficient to let us claim that people can treat the robot as a ‘social agent’ rather than a machine, because neither the ST nor the GOADI theory predicts how people react to social cues. As previously explained, both ToM and ST are at work in humans because ToM can
provide us insights for achieving successful social interaction in HHI that ST cannot fully explain (Williams et al., 2001). Therefore, we will look into what kind of social cues could influence people’s perception of the robot, and thus, perceive the robot as a ‘social agent’.

**Social cues in imitation**

According to (Holroyd, Rich, Sidner & Ponsler, 2011), to interact with people, the humanoid robot should have fluent and natural behaviors that are similar to, if not exactly like, human behaviors. In this section, we will turn to social cues that regulate HHI, which could be applied to HRI effectively.

In order to verify the GOADI theory in HRI, we need to make the robot present object-directed movements. In (Bekkering et al., 2000), the human model presented “use either hand to touch one of two dots on the table”; In (Wohlschläger et al., 2003), the human model presented “use either hand to put a pen into one of two cups” and made the imitator to follow the movement. For a human model, these movements are very easy. However, it’s very complex to make a robot model execute this movement. Therefore, we need to adjust this movement to make it easier for the robot to present. But before that, we need to step back to figure out how to effectively draw the imitator’s attention onto the object.

Drawing the imitator’s attention onto the object is very important in presenting object-directed movement. And this can be achieved by the ‘joint attention’ mechanism (Moore & Dunham, 1995) (as cited in Sugiyama, Kanda, Imai, Ishiguro, & Hagita, 2006), which refers to achieve shared focus of two individuals on an object by means of eye-gazing, pointing, or other verbal or non-verbal indications. For example, in HHI, in a dyadic conversation, the speaker can draw the listener’s attention by establishing eye contact with the listener, then looks at the target object to talk about it, pointing at the object while saying ‘look at this’ (Sugiyama et al., 2006). However, in a natural and smooth social interaction, the story usually doesn’t end here. Most likely, after the speaker said ‘look at this’, he/she will look at the listener again, to expect some reaction from the listener. And then, the initiatives of the speaker and the listener will exchange. And this process is referred to as turn-taking. According to (Ito & Tani, 2004), 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 can function as switching the initiatives in interactions among agents.
Therefore, in HHI, both the ‘joint attention’ mechanism and the ‘turn-taking’ are necessary to achieve a successful interaction regarding drawing one’s attention onto the object. In the following, we will turn to ‘joint attention’ mechanism and ‘turn-taking’ to regulate a smooth social interaction.

**Joint attention mechanism**

In imitation, the communication between two agents is mostly nonverbal. We can turn to the ‘joint attention’ mechanism to make the robot demonstrate the nonverbal social cues such as pointing and eye-gazing onto the object to draw the imitator’s attention. According to (Williams, Abidi, Gärdenfors, Wang, Kuipers & Johnston, 2013), drawing people’s attention to objects within perceptual vicinity will be a critically important ‘social skill’ for the robot. Moreover, attention is necessary for activation of an action representation (Castiello, as cited in Belopolsky, Olivers, Theeuwes, 2008). Williams *et al* (2013) implemented the joint-attention mechanism on a navigating robot and proved its effectiveness experimentally. They demonstrated that people can interpret robot pointing behavior in a similar way as human’s pointing behavior, and moreover, people tend to take gaze cues (the head trajectory) into consideration when they determine the robot’s pointing target.

**Turn-taking**

Turn-taking functions as initiating ‘joint attention’. As mentioned earlier, in the context of a dyadic conversation in HHI, turn-taking is an important ingredient whereby the role switch (the speaker, and the listener) is not determined by external sources but emerges from the interaction (Kose-Bagci, Dautenhahn & Nehaniv, 2008). Humans typically ‘know’ when to start and stop their turns in the social interactions in HHI, based on various factors like the context, the nonverbal feedback from the social interaction partners, such as gaze cues. 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 (Fiore *et al*, 2013). Kendon (1967) (as cited in Jokinen, 2010) was one of the first to emphasize gaze as a turn yielding and turn holding cue: he observed that the listener’s responses were quicker if there was a mutual gaze and there were delayed if the previous utterance terminated without a speaker gaze. Likewise, Gu & Badler (2006) pointed out that social gaze cues provide turn-taking signals to regulate the flow in a dyadic conversation.

The turn-taking cue indicated by gaze in a dyadic conversation is quite similar to that in a dyadic imitation task. Moreover, turn-taking is one of the particular social interactions in
imitation. In an imitation task, after the presenter (the robot) finished the action, it’s necessary to give cues or to draw attention, allowing the followers (people) to start imitation. There are many ways to draw attention in this context, such as using light indicator, or verbal communication like 'please imitate me' or just waiting. However, gaze cue provides a natural, smooth, and social way to regulate the flow in imitation. Many studies (Breazeal, 2000; Mutlu, Shiwa, Ishiguro & Hagita, 2009; Chao & Thomaz, 2010; Chao, Lee, Begum & Thomaz, 2011) found that smooth HRI benefits from turn-taking cues regulated by gaze.

The current study

In the current study, in the context of imitating a humanoid robot, we will look into whether the GOADI theory, which was initially tested in HHI, also applies to HRI. We want to verify whether people can imitate the goal of the robot’s action rather than the exact means of the robot’s action. However, in the GOADI theory, the object-oriented movement that the human presenter showed is difficult to implement on a humanoid robot. Thus, we turned to the ‘joint attention’ mechanism, we found that pointing, and gaze cues are very effective in HRI. And this should be easier and sufficient for a robot to present the object-oriented movement. As shown in Figure 1, we decide to let the robot present hand pointing with gaze pointing behavior towards the objects (two cups) to present an object-oriented movement. And similar to the initial GOADI experiment (i.e. use either hands touch one of two dots’ experiment), we decide to let the robot use its either hands, pointing either objects, with goal object present or absent.

It is hypothesized when the goal object is present, people tend to choose the correct object (correct refers to a mirrored fashion imitation); however, the means of achieving the object is often ignored. When the goal object is absent, people tend to pay more attention to the means of the movement. In other words, people will switch their hand use, their movement pathway more often in goal object present condition than that in goal object absent condition. But they
will choose the correct target more often in a goal object present condition than that goal object absent condition.

On the other hand, even though people can extract the goal from the robot’s action and imitate the robot similar to when imitating a human, it is not sufficient to claim that people can treat the robot as a social agent. Because the GOADI theory only provides us the motor aspects in imitation and it cannot predict how people react to social cues, we added a social cue. Specifically, we added the turn-taking cue indicated by eye-gaze, which can effectively regulate HHI. We wonder whether the turn-taking cues would affect the timing of imitation in HRI. We manipulate the gaze timing at which the robot looks at people with respect to its hand movement. We expect that the earlier the robot gazes at people, the RT of people to imitate will be shorter, but only if it is perceived as a social cue. However, we don’t expect that people will imitate differently from the GOADI suggested, because the GOADI theory didn’t give predictions on the social cues. In addition, we speculate that the reaction time of people to imitate will be shorter in a goal object present condition than that in goal object absent condition.
**METHOD**

**Task**

The experiment was conducted where participants were requested to play an imitation game with a humanoid Nao robot (Aldebaran Robotics, France). Participants needed to imitate the movement of the robot after the robot completed its action. The movement of the robot is either hand pointing to goal objects or without goal objects with four different timings of gaze cues. In total, there were eight conditions. In addition, for 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 pathways using trakSTAR (Ascension Technologies).

**Participants**

Participants were recruited using the participant database of the Human-Technology Interaction department of the Eindhoven University of Technology, the social network Facebook and the flyers sent 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 to 46). Among them, 68 participants were right-handed and 47 already had prior experience with the Nao robot. Participants received 7.5 euros as compensation or 9.5 euros if they were not affiliated with Eindhoven University of Technology.

**Experiment design**

The experiment was conducted as a two-factor within-subject design. 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 same cups were placed on the table, so that 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. A counterbalanced design was used for goal object factor. Therefore, half of the participants received goal-absent condition first, and then goal-present condition. The other half received goal-present condition first, and then goal-absent condition. In this way, order effects were removed when analyzing the effect of goal object factor.
The robot’s gaze timing was included as a second within-subject factor. The four gaze conditions are depicted in Figure 2. As shown in Figure 2, the robot hand movement time (MT) was 1 MT, including the time needed for moving back and forth, which lasted three seconds. During each trial, the robot looked at the participant (Figure 3a) before moving. When pointing, it looked at its hand pointing direction (Figure 3b), after some time, depending on the condition, it looked back at the participant (Figure 3a). Four different gaze timings at which the robot looked back at the participant with respect to its hand MT were implemented: (1) 0 MT: the robot always gazed at the participant, and it never gazed to its hand movement direction. (2) 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 seconds, 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. (3) 1 MT: After completing one hand movement, which was after 3 seconds, the robot looked back at the participant. (4) 1.5 MT: After one and a half MT, which was after 4.5 seconds, the robot looked back at the participant. This meant after the robot completing its hand movement, its gaze was still not back onto the participant, after another half of the hand MT, the robot looked back at the participant.

![Figure 2: Four gaze timing conditions](image-url)
Participants’ reaction time (RT) to initiate the imitation was introduced as 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 (left hand ipsi-lateral, left hand contra-lateral, right hand ipsi-lateral, and right hand contra-lateral movement). All the measurements were repeated for three times. In total, 2(Goals)*4(Gazes)*4(Movements)*3(repetitions), which resulting in 96 trials per participant.

Experiment setup

The experiment was carried out in the UseLab of the Eindhoven University of Technology. The humanoid Nao robot developed by Aldebaran Robotics (France) was used for the experiment. For giving clear instructions to participants and providing different types of movements for participants to imitate, the following characteristics of Nao were of importance: the voice synthesis and its 25 degrees of freedom which enabled us to control its head, limbs, or torso movement. In addition, the software Choregraphe was used to generate Nao’s body movements. While for voice instructions and other complex behaviors, python scripts were implemented using the Naoqi software development kit.

Figure 4 demonstrated a top-down view of the experiment setups. Participant sat face-to-face with the Nao robot. The experimenter observed participant’s performance through a see-through mirror behind the participant. The red imaginary line implied left-right symmetry. Two green dots represented two cups (goal objects) on the table. The distance from the center of the cup to both participant’s hands and robot hands was equal, which was 16cm. Two cups were 15.5cm respective from the red imaginary line, so 31cm apart. And this was the same...
for two grey crosses on the table. The radius of the grey cross was 3cm. The center of the grey cross (the black dot) indicated where participant’s index fingertips needed to be put.

The measurement device, the trakSTAR is a high-accuracy electromagnetic tracking device for short-range motion tracking applications (Ascension Technology Cooperation, 2011). It composes of a trakSTAR transmitter, two small sensors and electronics unit. The transmitter has a maximum range of 90cm. The sensor has an update rate of 100 HZ, and they were attached to participant’s index fingertips with skin tapes, so the movement pathways can be tracked in real-time. Because we need to deduce the real-time location (time information in unit of seconds; X, Y, and Z positional coordinates in unit of millimeters) of the sensors relative to the transmitter, it’s important to keep the transmitter in the same position for each participant. Likewise, it’s crucial to calibrate the trakSTAR system. The sensors were calibrated by instructing the participant to put their index fingertips on the black dot and setting the values for X, Y, Z coordinate for both hands to zero. Then we were able to collect accurate data of the movement pathways. The application MimicTrack was used for calibration. It is custom software developed by the HTI group of the Eindhoven University of
Technology. Using the timestamp information of both the robot and the trakSTAR, we extracted RT; it’s essential to use the same computer for recording data, otherwise comparing timestamps on unsynchronized clocks of different systems would be invalid.

**Manipulation of hand movement and gaze timing**

Gaze is a combination of both head and eye movement. However, the eye movement of Nao differed considerably from human visual system: Nao’s “eyes” are composed of two cameras with limited field of view and limited resolution. Moreover, the “eyes” cannot rotate within its head. Even though, Nao can still establish eye contact with participants by simply orienting its head. According to Cuijpers & van der Pol (2013), the perception of the gaze-direction and, specifically, eye contact with Nao as a looker, is similar compared to a human looker. Therefore, in this experiment, Nao’s head movement represented its gaze. The “gaze” can be decomposed into head yaw (i.e. left and right direction) and head pitch (i.e. up and down direction) as shown in Figure 5: x-axis represents time in seconds while y-axis represents head movement in degrees. Both hand movement and head movement started at 0.4s, which was served as a “buffering time” for Nao to start its movement stably and smoothly. All the slopes of starting movements and ending movements were the same except for the gaze timings.

![Figure 5: Four different gaze timings in seconds by head movement in degrees](image-url)
As depicted earlier in the experiment design section, there were four different gaze timings at which the robot looked back to the participant with respect to its hand movement time. All the starting and ending positions of Nao’s head and hand movement were the same for these four different gaze timings. The movements between the starting and ending positions depended on the combination of which type of hand movement Nao was using and which kind of head movement Nao was doing.

Nao’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 Figure 6: a. left hand ipsi-lateral movement. b. left hand contra-lateral movement. c. right hand ipsi-lateral movement. d. right hand contra-lateral movement. Note that the left and right are defined with respect to the participants’ view. Figure 7 depicted 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. Along with four different types of gaze timings, there were sixteen different gestures in total. Those sixteen gestures were repeated three times and pseudo-randomized (i.e. manually modified the order when the same gestures would appear repeatedly two or three times) for both goal-present and goal-absent condition.

![Figure 6: A front view of Nao’s hand movement](image1)

![Figure 7: A front view of the combination of Nao’s hand and gaze (head movement)](image2)
Measures

TrakSTAR was used for extracting participant’s RT, and hand movement pathways. In addition, which hand participants used to imitate, and which target participants chose was extracted. From this, which kind of imitation pattern (mirror-imitation or anti-mirror imitation) participants adopted was determined. In the following part, we explain in detail how the trakSTAR was used to extract this information from the raw data.

Data extraction using trakSTAR

In Figure 8, the timeline of each event was shown. The first row demonstrated a participant’s hand imitating event. The second row showed the robot’s hand movement event. And the third row depicted the trakSTAR marker event, which indicated when it triggered the robot movement. Now we look the events from the third row to the second row. After the participant’s hands were held for 3s within a range of 3cm radius from the starting position, the trakSTAR program wrote a marker ‘1’ to a file, which triggered the robot. The gesture of the robot hand movement lasted 3 seconds. Now we look at the events from the second row to the first row. After the completion of the robot’s gesture, the participant started to imitate (T2). When participants finished imitating, they put their hands back onto the starting positions. This process is repeated till the end of the experiment. The real-time movement pathways for both hands were recorded by the system with respect to the starting position (one for each hand). The timestamps, when the robot started doing its gesture (Trobot), was also recorded by the computer. Subsequently, a trakSTAR file including left and right hands
movement pathways, the corresponding timestamps, and markers were generated. Additionally, a condition file including all the conditions matching with 96 trials, and Trobot were generated. These raw data were further processed for extracting RT, movement pathways and additional information.

**Preliminary analysis**

MATLAB 2013a was used for the preliminary analysis and movement pathways extraction. First the trakSTAR data were split into segments using the marker ‘1’. The starting times (Trobot) of the robot movement were used to match the trial conditions to the movement recording. For example, in Figure 8, the matching process was done by making sure Trobot was between Tstart and Tend.

After the matching process, the displacement of left and right hand movement were calculated: 

\[ d_{left} = \sqrt{L_x^2 + L_y^2 + L_z^2} \quad d_{right} = \sqrt{R_x^2 + R_y^2 + R_z^2} \]

Then the maximum displacement of right and left hand were found and compared to determine which hand the participant chose to imitate the robot. If the maximum displacement of right hand was larger than that of left hand, we determined a right hand movement, otherwise a left hand movement.

After determining which hand participant used, we need to determine T2 for extracting RT. Assuming it was a right hand movement. It was obvious that T2 should be before the timestamp of right hand maximum displacement. After narrowing timestamp range, the mean and standard deviation of the first one hundred right hand displacement data were calculated. T2 was determined by a threshold: the mean (calculated from the first 1s data) + 2*standard deviation. To determine when the participant started to imitate (T2), 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.

**Movement pathways extraction**

Figure 9 depicted a left hand movement in the trakSTAR coordinate. The negative x axis represented forward movement, the positive y-axis represented leftward movement, and the positive z-axis represented downward movement. Figure 10 showed four typical movement pathways for a participant. The red circle indicated the starting position of the imitating hand. The grey cross indicated the approximate starting position of the other hand. The blue
trajectory showed the movement paths of the imitating hand. Green dots represented the approximate left and right target position. The coordinates of Figure 10 were adjusted in order to intuitively understand the movement paths from a top-view: **a.** left hand ipsi-lateral movement. **b.** right hand ipsi-lateral movement. **c.** left hand contra-lateral movement. **d.** right hand contra-lateral movement.

Figure 9: a left hand movement in the trakSTAR coordinates: negative x values represent forward movement, positive y values represent left ward movement, and negative z values represent upward movement.

![Figure 9: a left hand movement in the trakSTAR coordinates: negative x values represent forward movement, positive y values represent left ward movement, and negative z values represent upward movement.](image)

In Figure 10, the red star indicated the left maximum lateral displacement (LD), and the green star indicated the right maximum lateral displacement (RD). This information was used for...
extracting which target that the participant chose. In general, we compare LD and RD with the criteria ‘100 mm’ sideway displacement. For example, for a right hand movement, if both absolute value of LD and RD are less than 100 mm, then we determine it’s a right hand ipsilateral movement, in other words, the right target was chosen (as shown in Figure 10b). If LD is larger than 100 mm, and RD is less than 100 mm, then we determine it’s a right hand contra-lateral movement, in other words, the left target was chosen (as shown in Figure 10d).

Likewise, for a left hand movement, we can also use this way to extract which target that participant chose (as shown in Figure 10a & Figure 10c). However, each participant’s imitation pattern varies. For instance, the criteria ‘100 mm’ sideway displacement may not be suitable to every participant. For some participants, for a right hand movement, when the absolute value of LD exceeds 90 mm, it’s accounted for a left target chosen. Moreover, there are many possibilities of combinations with LD, RD and the criteria sideway displacement. For example, it could be both LD and RD are larger than 100 mm, which is difficult to determine which target was indeed chosen. Under these circumstances, we will change the ‘criteria’ within each participant, and combine with the visual inspection to extract the chosen target information.
**Procedure**

The participants were first guided to the UseLab and requested to sign the informed consent form after careful reading. Further explanations were given by the experimenter if participants were not clear about the procedure.

When there were no further questions, participants were asked to sit face-to-face with Nao as demonstrated in Figure 4. The experimenter adhered skin tapes to participants’ hands, and attached trakSTAR sensors to participants’ index fingertips. After that, the participants were asked to put their index fingertips onto the starting positions (shown as black dots in Figure 4). After checking whether all was set up correctly (i.e. goal objects had been put on or remove off the table, the skin tapes were in good adhesion, etc.), the experimenter started the experiment 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 two channels of trakSTAR (correspond to participant’s left and right hands). Then a five-trial training session was presented in order to let participants familiarize with the actions employed in the task. At the end of training session, participants were given one minute to relax. The purpose is also for Nao to relax to prevent its it from overheating. If participants asked the experimenter, such as, whether they need to use mirror imitation or anti-mirror imitation, the experimenter would suggest them doing the pattern they preferred.

The experiment consisted of two parts. The order of these two parts was counterbalanced. In the first part, the forty-eight trials were done with cups (goal-present) or without cups (goal-absent). After finished the first part, a four-minute relaxing were provided, for the Nao as well. In the meantime, the experimenter would remove or put cups onto the desk, then, followed up with the rest trials. Additionally, in the halfway of each part, the encouragement sentence “You are doing great! Keep moving” or “Good job! Keep moving” would be told by Nao. This is for the purpose of preventing participants’ boredom of the repeated task.
After finished the last trial, participants were asked to fill in their basic information, like age, gender, whether they were right handed or not, whether they had prior experience on Nao robot.

The whole experiment lasted about 35 minutes, and ended with a final debriefing to the participants. This experiment was conducted in accordance with the ethical standards laid down in the 1975 Declaration of Helsinki.

**Data analysis**

Prior to the statistical analysis, all erroneous trials that caused by trakSTAR double triggered the robot movements were deleted from the data set (8 out of 6720 trials (96 trials*70 participants), 0.12%). Additionally, all erroneous trials caused by participants were removed from the data set (12 out of 6712 trials; 0.18%). The erroneous trials were defined as the absolute standardized RT is larger than 4s. Moreover, one participant needed to be removed from the data, as in 76.4% trials, the participant adopted contra-lateral imitation. 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%.
**RESULTS**

In this section, the hand movement data and RT data gathered during the experiment was analyzed, and the corresponding results were reported. In the first part, we examined the hand movement data: first, the general imitation patterns of participants were divided into two groups. The follow-up analysis was performed separately for these two groups. Within each group, we examined specific imitation patterns that participants made, including hand switching, target switching, movement pathways switching, etc. Then we checked whether the goal manipulation had an effect on those specific imitation patterns respectively. In the second part, we examined the RT data: the main effect of goal, gaze timings, and the interaction effect of both were analyzed.

**HAND USE DATA**

‘Mirror symmetric’ Group and ‘anti-mirror symmetric’ Group

First of all, we examined which hand the participant used to imitate the robot’s movements. For example, when the robot used the hand on the right (left) side of the participant and the participant imitated with the right hand (left hand), we defined it as a ‘mirror symmetric’ imitation. However, when the participant imitated with left hand (right hand), we defined it as an ‘anti-mirror symmetric’ imitation. Note that the ‘left’ and ‘right’ were determined with respect to the participant’s view. Therefore, when the robot was doing a ‘right’ hand movement, he actually used his left hand with respect to his own body side.

![Figure 11](image_url)  
**Figure 11**: sum of ‘anti-mirror-symmetric’ imitation trials in both goal present and goal absent conditions of each participant. For each condition, there were 48 trials in total.
In Figure 11, the x-axis represented 70 participants’ ID, and the y-axis represented the sum of trials that they adopted ‘anti-mirror symmetric’ imitation. There are 48 trials in total in each goal absent and goal present condition. All participants mostly stick to their imitation ‘habit’ in both conditions. Participants didn’t switch their hand abruptly from one condition to the other condition, although there were some small amounts of hand switches. Overall, each participant’s hand use pattern could be clearly recognized: among 70 participants, 15 of them (ID: 1, 2, 15, 19, 21, 22, 23, 27, 31, 33, 41, 43, 59, 65, 69) (21.4%) adopted ‘anti-mirror symmetric’ imitation, while the rest adopted ‘mirror symmetric’ imitation. Two groups, ‘mirror symmetric’ (MS) group and ‘anti-mirror symmetric’ (AMS) group were divided according to the participants’ hand use.

The general imitation patterns by two groups

In Figure 13, the percentages of movements are shown for each movement of the robot and the participants. The movements made by the robot were on the left axis, and the movements made by the participants were on the right axis. For the MS group, the red pillars represented ‘correct’ imitation patterns (97.3%). Note that ‘correct’ refers to the chosen hand, the chosen goal (target) and the chosen movement pathway (ipsi-lateral or contra-lateral) are all correct in a mirror-symmetric fashion (as shown in Figure 12a). The grey pillars represented exact opposite imitation to the ‘correct’ imitation, which is ‘anti-mirror-symmetric’ imitation (1.1%). For the AMS group, the ‘correct’ imitation patterns were represented as the blue pillar (93.6%). Likewise, ‘correct’ refers to the hand, the goal object, and the movement pathway are all imitated in an anti-mirror-symmetric fashion (as shown in Figure 12b). The grey pillars represented that participants switched from ‘anti-mirror-symmetric’ pattern to ‘mirror symmetric’ imitation pattern (2.4%). Therefore, we reasoned that the mirror-symmetric imitation was easier than anti-mirror-symmetric imitation as the amount of ‘correct’ choices of the MS group is 97.3%, which is larger than 93.6% in the AMS group. For both groups, the green series (the light green and dark green pillars) represented the movement path errors that the participant generated (1.6% in MS group and 4.0% in AMS group). The movement path error is defined as: when the robot did a contra-lateral (ipsi-lateral) movement, participants responded with an ipsi-lateral (contra-lateral) movement. This type of error can be further divided: when the robot did contra-lateral movement (RC), and participants responded with ipsi-lateral imitation (PI), we defined it as RC-PI-error. Likewise, when the robot did ipsi-lateral movement (RI), and participants yielded contra-lateral imitation (PC), we defined it as RI-PC-error. In Figure 13, the dark green pillar represented
RC-PI-error (0.9% in MS group and 2% in AMS group). And the light green pillar represented RI-PC-error (0.7% in MS group and 2% in AMS group). The eight pillars which were surrounded in the yellow square represented the trials that participants switched their ‘correct’ imitation hand to the other hand.

Figure 12: a. The ‘correct’ mirror-symmetric imitation fashion. b. The ‘correct’ anti-mirror-symmetric imitation fashion. Blue arrows represent contra-lateral movement, and red arrows represent ipsi-lateral movement.

Figure 13: the general imitation pattern of both groups. The movements made by the robot are on the left axis, and the movements made by the participants are on the right axis. Red and blue pillars respectively represent the ‘correct’ imitation pattern for each group. Grey pillars represent the reversed imitation fashion. Green series represent the movement path errors that participants produced, of which dark green pillars represent RC-PI error, and light green pillars represent RI-PC-error. The yellow squares present the pillars that participants switched their ‘correct’ imitation hand to the other hand.
According to GOADI theory, the hierarchy for imitation is: when the goal is present, the chosen target is in the highest level, followed by the effector and/or the movement pathways. To interpret GOADI theory based on our data, we looked into the ‘error’ imitation that participants made in the goal-absent and the goal-present condition. The assumption is: participant made fewer errors for the chosen hand and the chosen movement path in the goal-absent condition than that in the goal-present condition. But participants will be more prone to choose the same target as the robot did in the goal present condition than in the goal absent condition. In the following section, we will examine whether the goal manipulation have an effect on the chosen hand, the chosen target and the chosen movement pathway, respectively for both MS and AMS groups. Last, we will examine the hierarchy of GOADI theory: whether participants made the least errors on the chosen target, followed by the chosen hand and/or the chosen movement pathway in a goal-present condition.

The effect of goal on hand switch

Hand switch is defined as the participants use the other hand instead of the ‘correct’ hand, and ‘correct’ is dependent on the MS or the AMS group. The assumption is that when a goal is present, participants tend to switch hands more often than that when a goal is absent. Figure 14 showed the percentages of participants’ hand switches.

In the MS group, when the goal is absent, there were 1.7% hand switches. And when the goal is present, there were 1.5% hand switches. An independent-samples t-test was run to determine if there were differences in hand switches in the goal-absent and the goal-present conditions. The result suggested that the hand switches in the MS group didn’t differ between conditions: $t(5168) = 0.784, p = 0.433$ (Table 1).

While in the AMS group, when the goal is absent, there were 1.4% hand switches. The value increased substantially to 4.2% in the goal-present condition. The result of independent-samples t-test indicated that the hand switches in the AMS group differed between conditions: $t(1182.094) = 3.033, p = 0.002$ (Table 1). Therefore, the goal manipulation has an effect on the chosen hand, but only for the AMS group.
The effect of robot movement and goal manipulation on hand switch

In Figure 13, the total amount that participants made hand errors was indicated in the yellow squares for the MS and the AMS group, respectively. Note that both the dark green pillar (RC-PI-error) and the light green pillar (RI-PC-error) were inside the yellow square. Thus we suspected that the hand switch could depend on whether the robot is doing a contra-lateral movement or an ipsi-lateral movement. As goal manipulation has an effect on hand switch in the AMS group, we inferred that there might be an interaction effect between goal manipulation and robot movement on participants’ hand switches.

A generalized linear model with binary logistic function was run to verify this assumption. The hand switch was made the dependent variable, while goal manipulation and the robot movement were made the fixed factors. Figure 15 showed the percentages of hand switches for both group on the effect of goal manipulation and robot movement.

In the MS group, the result showed that the ‘Hand switch’ percentages were 1.1% statistically significant more often in those trials that robot did contra-lateral movement than those of ipsi-lateral trials, with Wald $\chi^2(1) = 10.935$, $p = 0.001$. As for the goal manipulation, participants
switched hand 0.3% more often in the goal-absent condition than that in the goal-present condition. However, this difference is not statistically significant, with \( \chi^2(1) = 1.949, p = 0.163 \). And the interaction effect of the goal and the robot movement was close to significant: \( \chi^2(1) = 3.384, p = 0.066 \).

In the AMS group, in terms of the robot movement, ‘Hand switch’ were 0.1% more often in the robot ipsi-lateral trials than that in the robot contra-lateral trials, but not significantly so: \( \chi^2(1) = 0.858, p = 0.354 \). In terms of the goal manipulation, ‘Hand switch’ percentages were 2.7% statistically significant more often in the goal-present trials than the goal-absent trials, with \( \chi^2(1) = 8.630, p = 0.003 \). However, the interaction effect of the goal and the robot movement was not significant: \( \chi^2(1) = 2.684, p = 0.101 \).

![Figure 15: the effect of goal and robot movement on participant’s hand switch percentages](image)

**The effect of goal on the target switch**

After examining the effect of goal and robot movement on hand switch, we examined whether goal manipulation effected on the chosen target for both groups. In the MS group, when participants did a ‘correct’ imitation, they chose the same target as the robot. So our assumption for the MS group is: participants switch the chosen target less in the goal-present condition than that in the goal-absent condition. While for the AMS group, it was reversed, as the ‘correct’ imitation for them is to choose the different target as the robot did. When the target is present, we inferred that for the AMS group, participants were prone to choose the same target as the robot. In other words, for the AMS group, participants switch the chosen target more often in the goal-present condition than that in the goal-absent condition.
Figure 16 showed the percentages of target switch in two conditions. In the MS group, when a goal is absent, there were 2.9% target switches. And when a goal is present, there were only 1.6% target switches. An independent-samples t-test was run to determine if there were differences in target switch between the goal-absent and the goal-present conditions in the MS group. The result suggested that the target switch in the MS group did differ between conditions: \( t(4789.906) = 3.114, p = 0.002 \) (Table 2).

While in the AMS group, when the goal is absent, there were 4.6% target switches. When the goal is present, the target switch percentages were increased substantially to 7.4%. The result of independent-samples t-test indicated that the target switch in the AMS group differed between conditions: \( t(1362.173) = -2.252, p = 0.024 \) (Table 2). Therefore, the manipulation of goals has an effect on the target chosen for both groups.

![Figure 16: the effect of goal on the percentages of participants' target switches](image)

<table>
<thead>
<tr>
<th>Goal Absent</th>
<th>Goal Present</th>
<th>Effect of goal on target switch</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Mirror Group</td>
<td>0.029</td>
<td>0.167</td>
</tr>
<tr>
<td>Anti-Mirror Group</td>
<td>0.046</td>
<td>0.209</td>
</tr>
</tbody>
</table>

Table 2: An independent-samples t-test testing the effect of goal manipulation on participants’ target switches in two groups

The effect of goal on movement pathways

The movement pathways were defined as that participant doing an ipsi-lateral movement or a contra-lateral movement. The assumption is: For both groups, when a goal is absent, participant switched movement path less than that in the goal-present condition.
Figure 17 showed the percentage of movement path switches. In the MS group, when a goal is absent, there were 1.5% movement path switches. And when a goal is present, there were 1.6% movement path switches. An independent-samples t-test was run to determine whether the difference in movement path switch between the goal-absent and goal-present conditions is significant. The result suggested that the movement path switch produced in the MS group didn’t differ between conditions: $t(5168) = -0.223, p = 0.824$ (Table 3).

While in the AMS group, when a goal is absent, there were 3.1% movement path switches. When a goal is present, the movement path switches were increased to 4.9%. The result of independent-samples t-test indicated that the movement path switch in the AMS group also didn’t differ between conditions: $t(1361.771) = -1.778, p = 0.076$ (Table 3).

Although for both groups, the difference between conditions is not significant, the overall direction is in expectation: both groups switched movement path more in the goal-present condition than that in the goal-absent condition as shown in Figure 17.

![Figure 17: the effect of goal on the percentage of participants’ movement path switch](image)

<table>
<thead>
<tr>
<th></th>
<th>Goal Absent</th>
<th>Goal Present</th>
<th>Effect of goal on movement path</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Mirror Group</td>
<td>0.015</td>
<td>0.123</td>
<td>0.016</td>
</tr>
<tr>
<td>Anti-Mirror Group</td>
<td>0.031</td>
<td>0.172</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Table 3: An independent-samples t-test testing the effect of goal manipulation on movement path switch in two groups
**The RC-PI and RI-PC-error**

Though the movement path error did not differ significantly between conditions for both groups, however, the overall difference direction was expected. We further divided the movement path error into RC-PI-error (dark green pillar) and RI-PC-error (light green pillar) as depicted in Figure 13 in order to assess whether the goal manipulation has an effect on specific type of ‘RC-PI-error’ or ‘RI-PC-error’.

A generalized linear model with binary logistic function was run to verify whether the goal manipulation has an effect on specific type of ‘RC-PI-error’ or ‘RI-PC-error’. The movement path switch was made the dependent variable, while the goal manipulation and the robot movement were made the fixed factors. Note that the dependent variable “movement path switch” is actually referred to two kinds of errors dependent on which movement robot did. For example, when the robot did a contra-lateral movement, the movement path switch refers specifically to RC-PI error. Likewise, when the robot did an ipsi-lateral movement, the movement path switch refers specifically to RI-PC error. Figure 18 showed the percentages of movement path switch for both group on the effect of the goal manipulation and the robot movement.

In the MS group, the result showed that the ‘movement path switch’ percentages were 0.5% more in those trials that the robot did contra-lateral movement than those of ipsi-lateral trials, with Wald $\chi^2(1) = 1.799, p = 0.180$. Therefore, in other words, RC-PI error is 0.5% more than that RI-PC error in the MS group. As for the goal manipulation, participants switched 0.1% more movement path in the goal-present condition than that in the goal-absent condition.
However, this difference is not statistically significant, with $\chi^2(1) = 0.102$, $p = 0.749$. Nevertheless, the interaction effect of the goal and the robot movement was statistically significant: $\chi^2(1) = 11.999$, $p = 0.001$.

In the AMS group, in terms of the robot movement, ‘movement path switch’ was 0.2% more often in the robot ipsi-lateral movement than that of the robot contra-lateral movement, but not significantly so: $\chi^2(1) = 0.058$, $p = 0.810$. In other words, RI-PC error is 0.2% more than that RC-PI error. In terms of the goal manipulation, ‘movement path switch’ percentage was 1.8% more in the goal-present trials than that in the goal-absent trials, with $\chi^2(1) = 3.319$, $p = 0.076$. However, the interaction effect of goal and robot movement was not significant: $\chi^2(1) = 0.199$, $p = 0.656$.

Overall, the presence of the goal object seems to attract the participants to make RC-PI errors in the MS group. Likewise, in the AMS group, it seems that the presence of goal object attracts participants to make both RC-PI and RI-PC errors. Therefore, we will further assess in the goal-present condition, when participants made a RC-PI or RI-PC error, whether or not it is because they are prone to choose the same goal object as the robot chose.

First, we look into how the RC-PI or RI-PC-error is produced. Note that the errors are independent of which imitation ‘habit’ (Mirror-symmetric or anti-mirror-symmetric) the participants adopted. In Figure 19, the robot and the participant’s hand movement are shown. Blue arrows represent contra-lateral movements, and red arrows represent ipsi-lateral movements. There are two possible situations for producing a RC-PI or RI-PC-error as depicted in Figure 19. Because the goal is in the highest hierarchy, we reasoned that the probability of situation 2 (shown as S2 in green in Figure 19), which is the participant chose the same target as the robot, is larger than that in situation 1 (shown as S1 in light olive green in Figure 19) when making a RC-PI or RI-PC error.
Figure 19: two possibilities of making a RC-PI or RI-PC error. The blue arrows represent contra-lateral movements, and the red arrows represent ipsi-lateral movements. S2 in green represents that the participants chose the same target as the robot chose, while S1 in light olive green represents that the participants chose the different target as the robot chose.

There are in total 32 trials (MS group), and 18 trials (AMS group) that participants produced RC-PI-error in the goal-present condition. 16 out of 32 trials (50% of the MS group) and 14 out of 18 trials (78% of the AMS group), participants chose the same target as the robot did. The ratios are shown in the first row of Figure 20.

Figure 20: when the participants made RC-PI errors or RI-PC errors in a goal present condition, the ratio that they chose the same target as the robot to the different target.
Likewise, there are in total 10 trials (MS group), and 17 trials (AMS group) that participants produced RI-PC-error in goal-present condition. 3 out of 10 (30% of the MS group) and 15 out of 17 trials (88% of the AMS group), participants chose the same target as the robot did. The ratios are shown in the second row of Figure 20.

From Figure 20, it is clear that in the AMS group, participants chose the same target as the robot chose more often than the other target when making both RC-PI and RI-PC error. Whereas in the MS group, they chose the different target from the robot more when making RI-PC-errors.

**The hierarchy of imitation**

As concluded by the GOADI theory, when a goal is present, the chosen target is in the highest level, followed by the effector and/or the movement pathways. To interpret this in a meaningful way on the data, it means participant produce the least errors on the chosen target, followed by the chosen hand and/or the chosen movement path in the goal-present condition.

Figure 21 demonstrated the percentages of errors for choosing the target, the hand, and the movement path in the goal-present condition. It’s clear that the AMS group made more errors than that in the MS group.

In the MS group, when a goal is present, the percentages of errors for choosing the target, the hand, the side, is 1.6%, 1.4%, and 1.6%, respectively (see Table 4). It turned out that the chosen hand is in the highest hierarchy (the lease error made), followed by the chosen target or the chosen movement path.

In the AMS group, the result is 7.4%, 4.2%, and 4.9%, respectively (see Table 4). It seems that the chosen hand is in the highest hierarchy, followed by the chosen movement path, and the chosen target seems to be in the lowest level.

<table>
<thead>
<tr>
<th>The percentage of errors in the goal-present condition</th>
<th>Goal</th>
<th>Hand</th>
<th>Movement path</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS group</td>
<td>1.6%</td>
<td>1.4%</td>
<td>1.6%</td>
</tr>
<tr>
<td>AMS group</td>
<td>7.4%</td>
<td>4.2%</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

*Table 4: the percentages of goal errors, hand errors, and movement path errors*
Figure 21: the percentages of goal errors, hand errors and the movement path errors in the goal-present condition.
RT DATA

The effect of goal and gaze timing

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

First, we presented the main effect of gaze timings. The result was shown in Figure 22 for the MS and the AMS group, respectively. Note that the robot movement time (3s) is initially included in the mean of RT. In order to make the scale of RT more meaningful, we subtracted 3s from the initial mean of RT.

In the MS group, Mauchly's test of sphericity indicated that the assumption of sphericity had 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 = 4.37, SD = 1.15) to gaze timing 3s (M = 3.72, SD = 0.73) to gaze timing 1.5s (M = 3.47, SD = 0.63). However, for gaze timing 0s (M = 3.59, SD = 0.59), the decreasing pattern didn’t continue. Post hoc analysis with a Bonferroni adjustment revealed that RT differed significantly within each gaze timing, though gaze timing 0s and gaze timing 1.5s differed in an opposite direction (in Table 5).

<table>
<thead>
<tr>
<th>Gaze 0 – Gaze 1.5</th>
<th>0.109</th>
<th>0.025</th>
<th>&lt; .001</th>
<th>[0.039, 0.178]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze 1.5 - Gaze 3</td>
<td>-0.244</td>
<td>0.034</td>
<td>&lt; .001</td>
<td>[-0.337, -0.152]</td>
</tr>
<tr>
<td>Gaze 3 – Gaze 4.5</td>
<td>-0.652</td>
<td>0.074</td>
<td>&lt; .001</td>
<td>[-0.856, -0.449]</td>
</tr>
</tbody>
</table>

Table 5: the mean difference within gaze timing for the MS group
Figure 22: the effect of gaze timings on RTs, error bars represented standard error mean. RTs were subtracted 3s of the robot movement from the initial RTs.

In the AMS group, Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(5) = 10.235, p = .07$. The gaze timing elicited statistically significant changes in RT, $F(3, 36) = 14.119, p < .0005$. Specifically, RT decreased from gaze timing 4.5s ($M = 4.21, SD = 0.84$) to gaze timing 3s ($M = 3.89, SD = 0.63$) to gaze timing 0.5s ($M = 3.68, SD = 0.64$) to gaze timing 0s ($M = 3.60, SD = 0.59$). Post hoc analysis with a Bonferroni adjustment revealed that RT was not significantly differed within each gaze timing, except for gaze timing 0s to gaze timing 3s (mean difference = -0.277, 95%CI [-0.483, -0.075], $p = 0.006$), gaze timing 0s to gaze timing 4.5s (mean difference = -0.601, 95%CI [-1, -0.204], $p = 0.003$), gaze timing 1.5s to gaze timing 4.5s (mean difference = -0.528, 95%CI [-0.962, -0.095], $p = 0.014$).

Second, the main effect of the goal was analyzed. However, it didn’t show a statistically difference between goal conditions for both groups. In the MS group, RT decreased from the goal-absent ($M = 3.81, SD = 0.74$) to the goal-present ($M = 3.77, SD = 0.65$), with $F(1, 51) = 0.5, p = 0.483$. In the AMS group, RT increased from the goal-absent ($M = 3.82, SD = 0.65$) to the goal-present ($M = 3.87, SD = 0.68$), with $F(1, 12) = 0.274, p = 0.610$.

Last, we examined the interaction effect. However, there was no statistically significant interaction in RT between the goal manipulation and gaze timings for both groups (MS group (see Figure 23): $F(2.152, 109.774) = 1.697, p = 0.186$; AMS group (see Figure 24): $F(3, 36) = 0.487, p= 0.693$.).
Figure 23: the interaction effect of goal and gaze in the MS group, error bar represent standard error. RTs were subtracted 3s of the robot movement time from the initial RTs.

Figure 24: the interaction effect of goal and gaze in the AMS group, error bar represented standard error. RTs were subtracted 3s of the robot movement time from the initial RTs.
DISCUSSION

In the current study, we turn to imitation as a substrate for learning whether people can interpret the humanoid robot’s ‘mind’ as human-beings. We first investigated whether the GOADI theory, which was initially tested on HHI, was also applicable to HRI. Secondly, we looked into whether the social turn-taking cue indicated by eye gaze, was effective in HRI. By studying these two aspects, we can gain further insights about what kinds of behaviors can facilitate people to treat the humanoid robot as a ‘social agent’, thus gaining a smooth and successful HRI. The results turned out that the GOADI theory was partly in support of HRI, and the social turn-taking cue indicated by different eye gaze timings has a substantial effect on participants’ RT to initiate their imitation. In the following sections, we discuss our findings respectively for these two aspects, pointing out the limitations, suggesting possible implications for further research in the field of HRI, and drawing conclusions.

General discussion

The MS and the AMS imitation

In the studies of Bekkering et al (2000) and Wohlschläger et al (2003), only the MS imitation was taken into account. In the current study, we included both the MS and the AMS imitation to verify whether the GOADI theory is applicable in HRI. In general, there were two types of imitation patterns, the mirror-symmetric (MS) imitation and the anti-mirror-symmetric (AMS) imitation. Most participants (78.6%) preferred MS imitation. This is in accordance with the findings by (Bekkering et al, 2000; Wohlschläger & Bekkering, 2002; Perra & Gattis, 2008) that humans tend to imitate in a MS fashion rather than in an AMS fashion. Chiavarino, Apperly & Humphreys (2007) indicated that this preference could be explained by stimulus-response compatibility (SRC). In a MS (meaning that left and right stay the same, but forward and backward are reversed) imitation, the frame of the robot’s movements is mirrored to that of the participant. However, in an AMS imitation, the frame of the robot’s movements must be rotated to map onto the actions of the participant. Therefore, the MS imitation is simply easier to represent and to remember than the AMS imitation because the AMS imitation involves mental rotation (Chiavarino et al, 2007). Based on our data, this explanation is reasonable because we found that the participants in the AMS group made errors in 6.4% trials, which is more than the MS group (2.7%). Furthermore, in the MS group, only in 1.1% trials, the participants switched to the AMS imitation. While in the AMS group, the participants switched to MS imitation in 2.4% trials. Koski, Iacoboni, Dubea, Woods, &
Mazziotta (2003) indicated that the preference for MS imitation is because there is a reliably greater activity in human mirror neuron system for the MS imitation than that for the AMS imitation. Although this potentially explained why the MS imitation is easier than the AMS imitation and mostly adopted among participants, this view by Koski et al (2003) is to some extent consistent with the direct mapping view, which refers to the proposal that observed actions are automatically mapped from the visible movements of another to the proprioceptive controlled movements of the self (Perra & Gattis, 2008). However, we cannot directly turn to this direct perceptual-motor mapping view to interpret our findings. First, it was the robot’s movement that needed to be mapped, which means that it’s not that much ‘direct’ mapping because the robot has different limbs, body size as humans. Second, Belopolsky et al (2008) mentioned the ‘anatomical congruent’ representation. And it refers to, for example, the participant’s right hand corresponding to the robot’s right hand, which is on the left side of the participant. They claimed that if the representation of the observed action is created through the direct-matching onto the participants’ internal motor representation, then the anatomically congruent representation of the robot should be activated. If so, we would expect that the AMS imitation would be much easier than the MS imitation. However, this is not the case based on our result, and on other findings (Avikainen, Wohlschläger, Liuhanen, Hänninen, Hari, 2003; Chiavarino et al, 2007), which showed that the MS imitation is less error-prone than the AMS imitation and mostly adopted among participants. Moreover, Belopolsky et al (2008) indicated that in the study of (Koski et al, 2003), though the MS imitation has greater activity in mirror neuron system than the AMS imitation, in an observation-only condition, the reverse pattern of greater activity for the AMS representation rather than the MS representation was found. Thus, they speculated that there might be a transformation of the representation for imitation tasks in order to accommodate the goals of imitative action. Cuijpers et al (2006) indicated that action goals are not view-point dependent, which has substantial advantage over simply ‘direct perceptual-motor mapping’.

Bekkering et al (2000) and Perra & Gattis (2008) proposed a mediated mapping view, which holds that the mapping between perception and action during imitation is mediated by cognitive processes. According to them, the GOADI theory (Bekkering et al, 2000; Wohlschläger et al, 2003; Perra & Gattis, 2008) is the most articulated proposal for mediated mapping. As demonstrated in the GOADI theory, imitation depends on a hierarchical representation of action, based on its component goals, maintained in working memory. Only
the more important goal will be reproduced if the working memory resources are limited. If reducing the number of possible goals (objects), the imitation errors will reduce. And the difference in accuracy is due to differences in working memory capacity. Consequently, the preference for the MS imitation can also be explained by the GOADI theory because the AMS imitation might pose higher working memory demands on the imitator (Chiavarino et al, 2007).

The effect of the goal object

As illustrated in the theory of GOADI, the mapping of perception to action underlying imitation involves decomposing an observed action into elements and then reconstructing (Perra & Gattis, 2008). The decomposed elements in the current study are divided into three elements: the effector (participants’ hands), the goal object (two cups), and the movement path (ipsi-lateral/contra-lateral movement). Additionally, according to the GOADI theory, reducing the number of goals will reduce the imitation errors. Therefore, we will discuss whether reducing the goal (goal-absent condition) will reduce the imitation errors on the chosen target (goal error), the chosen movement path (movement path error), and the chosen hand (hand error) compared with the goal-present condition in both the MS and the AMS group, respectively. Last, we will discuss the hierarchy representation of imitation in these two groups.

The MS group

For the MS group, an ‘error’ means that participants didn’t imitate in an exact ‘correct’ MS way, generally, it is ‘goal error’, ‘movement path error’ and ‘hand error’.

In terms of the ‘goal error’, in 2.2% trials, participants made ‘goal error’. we expected that the participants made less goal error in a goal present condition. Indeed, the percentage of goal errors was significantly higher in the goal absent condition than that in the goal present condition.

In terms of ‘movement path error’, in 1.6% trials, participants made ‘movement path error’. We expected that without a goal object, the participants imitate the movement more closely. In other words, participants were expected to make less ‘movement path error’ in a goal absent condition. It turned out that the percentage of ‘movement path error’ was not significant differed between conditions.
In terms of ‘hand error’, we expected that people made more ‘hand error’ in goal present condition than that in the goal absent condition. However, the result showed there was no significant difference between the goal absent and the goal present condition. Therefore, for those three types of ‘errors’, only the ‘goal error’ differed between conditions, while the ‘movement path error’ and the ‘hand error’ didn’t differ between conditions.

We looked into some specific types of errors. First, we looked into the factor that might influence the hand error. From (Figure 13), it’s clear that the robot movement is one of the factors. And it turned out that people switched their hand significantly more often in the trials when the robot is doing contra-lateral movement (RC) than that when the robot is doing ipsi-lateral movement (RI). Note that choosing the wrong hand in the RC trial and RI trial also has the consequence that the path of the movement is incorrect (e.g. for a RC trial, participant produce ipsi-lateral movement instead of contra-lateral movement). As imitating a contra-lateral movement is more effortful than the ipsi-lateral movement, we assume that the participants might switch to ipsi-lateral movement in a RC trial more often compared with switching to contra-lateral movement in a RI trial. And this is expected as when the goal is present, it turned out that the participants produced 1.7% more RC-PI error than the RI-PC error. Although the effect is small, the direction is expected. Furthermore, the interaction effect of the robot movement and the goal manipulation is significant. When the goal is absent, participants made less movement path errors when the robot did contra-lateral movement (i.e. RC-PI error is less in the goal absent condition). While in goal present condition, people make less movement path errors when the robot did ipsi-lateral movement (i.e. RI-PC error is less in the goal present condition). The former is in an expected direction, because as suggested in the GOADI theory, people imitate means more closely when the goal is absent. However, the latter is a reversed pattern: more RI-PC error was produced when goal object is absent. It’s not clear what mechanism might underlie it.

Furthermore, we expect that when producing a RC-PI or RI-PC error, the participants would move towards the same target as the robot because the goal is very salient as GOADI indicated. However, contrary to our expectation, we found that in 50% of the goal-present trials, the participants made RC-PI error towards the same target as the robot, and in only 30% of the trials, the participants made RI-PC error towards to the same target as the robot. So it doesn’t seem the presence of the goal caused the participants to make RC-PI or RI-PC errors, rather, they prefer to use the ‘correct’ hand to choose the other target. Therefore, we
inferred that it’s the robot’s hand rather than the goal object to drive the participants to imitate.

Consequently, in the MS group, we found out that the imitation error decreased when there was no goal present, only for the ‘RC-PI error’. In other words, only when the robot did contra-lateral movement, participants imitate the movement path more closely when the goal is absent. However, the hierarchical organization in the GOADI theory is not observed as in the studies of Berkkering et al (2000) and Wohlschlager et al (2003). It turns out that the chosen hand is in the highest hierarchy, followed by the chosen target or the movement path (see Figure 21). And this is in accordance with the inference that the effector rather than the goal object drives imitation.

The AMS group

For the AMS group, an ‘error’ means that participants didn’t imitate in a ‘correct’ AMS way, generally, it is ‘goal error’, ‘movement path error’, and the ‘hand error’.

In terms of the ‘goal error’, in 6% trials, participants made ‘goal error’. (Note that the ‘goal error’ in AMS group is exactly ‘goal correct’ in a MS imitation.). We expected that participants made more ‘goal error’ in the goal present condition. Indeed, participants in the goal-present condition made 7.4% ‘goal error’, while in the goal-absent condition, they made 4.6% goal error. And this difference between conditions is significant. Thus, we might infer that the presence of the goal inhibited the participants to choose the ‘correct’ goal. In other words, the participants in the AMS group are more attracted to the same target as the robot chose, rather than the ‘correct’ target they needed to choose.

In terms of the ‘movement path error’, participants made these errors in 4% trials. And we expected that without the goal object, participants imitate more closely. Indeed, without a goal object, participants made 1.8% fewer ‘movement path error’, though not significantly so.

In terms of the ‘hand error’, in 2.9% trials, participants made ‘hand error’. Moreover, the ‘hand error’ decreased significantly when there was no goal present. This is as expected because without the goal, participants focus more accurately on using the ‘correct’ hand as the GOADI suggested. However, unlike the MS group, the participants made 0.1% more chosen-hand error in RI trial than that in RC trial. Specifically, participants made 0.2% more RI-PC error than RC-PI error. This is a reversed pattern as we found in the MS group. Counterintuitively, it seems the participants rather prefer the more effortful contra-lateral
movement than the easier ipsi-lateral movement. But thinking about the findings that the participants are more attracted to the target that the robot pointed but not the ‘correct’ target in an AMS sense that they need to choose, we reasoned that the RI-PC or the RC-PI error is made because: (1) the target is present; (2) the target that the robot pointed is more attractive than the other ‘correct’ target they need to choose. The evidence of (1) lies in when the goal object was absent, there were 2.2% less RC-PI errors and 1.5% less RI-PC error, though not significantly so. The evidence of (2) lies in (Figure 20): in 78% cases when the participants making the RC-PI error and in 88% cases when they making the RI-PC error, they pointed to the same target as the robot did.

Consequently, in the AMS group, we found out that the imitation error decreased when there was no goal present, for both the RC-PI and RI-PC error. This is in accordance with the GOADI theory, when the goal is absent, people imitate more closely of the movement path. Furthermore, the participants made least errors on the chosen hand, followed by the chosen movement path, and the chosen target is the most error-prone aspect. In Bekkering et al (2000) and Wohlschläger et al (2003), they only investigated the MS imitation, and they claimed that the goal object is in the highest hierarchy because people produced the least errors on the chosen target. However, we argued that in an AMS imitation, the least-errors-produced aspect doesn’t necessarily mean that aspect is in the highest hierarchy. Rather, the fact that the participants were attracted to the same target as the robot exactly reflected that the goal object is a very salient aspect, and therefore it is in the highest hierarchy in an AMS imitation.

The potential explanations for the hierarchy in imitation

Overall, based on our data, in the MS group, the GOADI theory only partly applies. When the goal is absent, people imitate more closely of the movement path, but the effector instead of the goal object is in the highest hierarchy. However, for the AMS group, although the least-produced errors are the chosen effector, we demonstrated that the goal object is a very salient aspect for people to imitate. In the following part, we presented possible explanations for those different hierarchies that we found in HRI.
The potential explanation is that we used shared goals (two cups) instead of proprietary goals (4 cups) between the robot and the participants (similar to the situations shown in Figure 25).

In the study of Perra & Gattis (2008), although they were able to demonstrate that the GOADI theory applies in shared goals condition in a MS fashion, they implied a few findings: “reducing the mapping between perception and action by eliminating the need to map the relation between the model’s set of objects and the imitator’s set of objects would reduce the demands on cognitive resources, and thereby facilitate mapping and enable the imitator to imitate other goals, which leads to fewer errors and thus improved imitation performance. In our study, the hierarchy of the MS group is not very clear. As shown in Figure 21, the condition for the MS group is approximately constant; participants produced similar percentages of errors, 1.6%, 1.4%, and 1.6% on the chosen goal, the chosen hand, and the chosen movement path, respectively. In addition, the percentages of errors produced in MS group are less than that in an AMS group, so the MS imitation is easier. And this could be enhanced since we used a shared-objects condition. Therefore, it is very difficult to detect any differences between the ‘goal error’, ‘hand error’, and the ‘movement path error’ in the MS group.

On the other hand, for the AMS group, the percentages of errors produced are larger for the ‘goal error’ than for the other error types. As mentioned previously, the AMS imitation might pose higher working memory demands on the imitator, which leads to the relatively ‘bad’ performance pattern for each error type.

**The effect of goal on RT**

For the effect of goal on the RT to initiate imitation, we speculate that when the goal is present, the participants will react faster. However, for both MS and AMS group, we didn’t
find any significant difference of the RT between the goal-present and the goal-absent condition.

In our experiment, most participants indeed paid more attention to imitate the hand movement instead of the goal object. This was concluded from the hand movement data, which we found that the hand use is in the highest hierarchy in imitation. Moreover, from the visual inspection from the trajectory data (see Figure 10), most participants imitated the exact movement path as the robot presented (i.e. for imitating a contra-lateral movement, they didn’t go directly as a straight line towards the cup, but, the trajectory is more like a “C” shaped as the robot presented). Thus, we speculate that the presence of the goal object didn’t help the participants to react faster. And this could lead to the no significant difference of the RT between the goal present and the goal absent condition.

The effect of turn-taking on RT

The result showed that the turn-taking indicated by eye gaze cue is quite effective. Generally, the participants’ RT to start to imitate is increased with the gaze timing. The earlier the robot looked at the participants with respect to its movement, the earlier the participants start to imitate. However, the ‘gaze timing 0s’ is a special case, because the robot never looked away in this condition. Thus, the gaze time is not well-defined, although we arbitrarily set it to 0s. Additionally, the turn-taking cue is not dependent on the manipulation of goal object, as there’s no interaction effect of these two variables. In spite of this, eye gaze is an effective social turn-taking cue.

Limitations and future work

A first limitation is that even though the Nao robot is very developed, there are still some constraints to make it to do the normal pointing movement. The Nao robot only has three fingers, and they cannot move independently. It cannot point like a normal human hand, with the index finger stretched, and with the other fingers closed. Instead, when the Nao is doing pointing movement towards the cup, it’s more like that the Nao intends to grasp and hold the cup, which is reflected in the observations of participants’ imitation performance.

A second limitation is that we used shared objects condition instead of the proprietary object condition, which makes the task easier. In the future work, propriety object condition can be used in verifying the GOADI theory in HRI, which could facilitate the participants to make more errors and perhaps show a clearer pattern of the hierarchy in the MS imitation.
Additionally, turn-taking is a highly multimodal process in HHI and it is the fundamental way that humans organize interactions with each other (Chao et al., 2011). For example, in HHI, we might use our verbal language, body language and combine with the gaze cues to inform the other person(s) that it’s their turn to do something. In the current study, gaze is the only indicator of turn change. However, in complex HHI, gaze alone is not sufficient to model all turn-taking behavior. For instance, in a multi-party conversation, we might have misunderstandings of turn-taking by merely gaze cue, under this circumstance, verbal language or body language will be very useful to repair this misunderstanding. Thus, future work can include enhancing social components to implement turn-taking.

Last, the data we gathered is very rich and informative. There are quite a few interesting aspects that we can further investigate. For example, although we inform the participants to imitate after the robot finishes its movement; some participants start to imitate the robot spontaneously when the robot starts its movement. We can imagine that the spontaneous imitation requires the participants to react to imitate in a really short time, and they need to extract the goal very quickly. We can look specifically into this group, to see whether they still imitate as the same as the GOADI theory suggested.

**Conclusions**

In the current study, we turned to imitation as a substrate for learning whether people can treat the humanoid Nao robot as a ‘social agent’. Specifically, we tried to verify the GOADI theory, which was initially applied to HHI, was also applied to HRI. This provide us some insights about whether human can imitate a robot’s movement as the same as when imitating the movements of human. However, the GOADI theory did not provide us with insights about whether people treat robots in a social way. Thus, we added the social turn-taking cue indicated by eye gaze. By studying these two aspects, we learned about what kinds of behaviors can facilitate people to treat the humanoid robot as a ‘social agent’, thus gaining a smooth and successful HRI.

The results turned out that the GOADI theory was partly in support of HRI. When the goal object is absent, the participants who imitated the robot in a MS fashion improved imitation performance, even though they mainly paid attention to imitate the chosen hand of the robot. It is likely that it is not the shared goal object driving them to imitate, rather, the robot’s hand is driving them to imitate. Meanwhile, when the goal object is absent, the participants who imitated the robot in an AMS fashion also improved their imitation performance. Although
they also made least errors on the chosen hand, they have been attracted to the same shared
goal object as the robot. Therefore, we reasoned it is the goal object rather than the chosen
hand driving the participants to imitate in an AMS fashion.

In the current experiment, most participants adopted the MS imitation. They paid more
attention to the chosen hand of the robot. Therefore we reason that they used a quite low level
representation rather than the high level goal representation to imitate. This could be caused
by the premature hand movement of the humanoid robot. Human hand is very advanced and
sophisticated, it can manipulate object, doing a lot meaningful gestures, etc. Therefore, in
designing a humanoid robot, not only its appearance need to be emphasized, but also, the
features and the functions of the hand, so that proper pointing movements can be made.

On the other hand, the human social behavior of turn-taking indicated by eye gaze is very
effective when implementing to the humanoid robot. The participants’ reaction time to
initiate imitation is decreased with the gaze timing with respect to the hand movement
decreased. This is a quite informative finding that we could use in other field of HRI
researches that related to turn-taking, such as multi-party conversation in HRI.

To wrap-up, with the current study, we gained further insights regarding the GOADI theory
and social aspects in HRI. We argue that to let people treat the robot as a potential ‘social
agent’, not only the social cue implemented is important, but also we need to focus on
designing more advanced hand movement that let people have a high level representation of
its movement.
REFERENCE


